

Capitolo 1

Equazioni alle Derivate Parziali

1.1 Introduzione

Un'Equazione alle Derivate Parziali (PDE) è un'equazione che mette in relazione una funzione incognita dipendente da due (o più) variabili indipendenti alle sue derivate parziali rispetto a queste variabili. La necessità di utilizzare tali equazioni sta nel fatto che queste consentono di descrivere in modo più accurato determinati fenomeni naturali. Infatti quando si cerca di descrivere fenomeni dipendenti da diverse variabili indipendenti (più comunemente posizione e tempo) allora è necessario utilizzare un modello differenziale alle derivate parziali. Un esempio di PDE è il seguente

$$a \frac{\partial^2 u}{\partial x^2} + 2b \frac{\partial^2 u}{\partial x \partial y} + c \frac{\partial^2 u}{\partial y^2} + f \left(x, y, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y} \right) = 0$$

dove anche a, b, c possono essere funzioni di x, y, u e delle derivate parziali prime di u . Generalmente le derivate parziali del secondo ordine possono essere indicate anche in forma più compatta:

$$u_{xx} = \frac{\partial^2 u}{\partial x^2}, \quad u_{xy} = \frac{\partial^2 u}{\partial y \partial x}, \quad u_{yy} = \frac{\partial^2 u}{\partial y^2}$$

e, analogamente quelle del primo ordine.

$$u_x = \frac{\partial u}{\partial x}, \quad u_y = \frac{\partial u}{\partial y}$$

cosicchè la forma dell'equazione appena vista diviene:

$$au_{xx} + 2bu_{xy} + cu_{yy} + f(x, y, u, u_x, u_y) = 0.$$

Le equazioni alle derivate parziali **del secondo ordine**:

$$a(x, y, u)u_{xx} + 2b(x, y, u)u_{xy} + c(x, y, u)u_{yy} + f(x, y, u, u_x, u_y) = 0$$

sono le più diffuse. Un'equazione alle derivate parziali si dice di ordine p se p è il massimo ordine di derivata che vi compare.

Generalmente la scelta delle variabili indipendenti dipende dal problema fisico: infatti le variabili x, y, z indicano una posizione nello spazio, mentre la variabile t indica il tempo. Talvolta anche le variabili x_1, x_2, x_3 indichino una posizione nello spazio. Considerando quindi le due equazioni

$$u_{xx} + u_{yy} + f(x, y, u) = 0, \quad u_{xx} + u_{tt} + f(x, t, u) = 0$$

esse sono matematicamente equivalenti ma fisicamente no, perchè la prima descrive un fenomeno indipendente dal tempo (cioè stazionario) che riguarda un dominio bidimensionale (la posizione è descritta dalle variabili (x, y)) mentre nel secondo caso il fenomeno descritto evolve nel tempo in un dominio monodimensionale.

Nell'equazione del secondo ordine

$$a(x, y, u)u_{xx} + 2b(x, y, u)u_{xy} + c(x, y, u)u_{yy} + f(x, y, u, u_x, u_y) = 0$$

non compare la derivata u_{yx} perchè, in ipotesi di continuità, applicando il Teorema di Schwarz, le derivate parziali miste sono uguali, cioè:

$$u_{xy} = u_{yx}.$$

Va anche precisato che nelle equazioni più diffuse non è presente la derivata u_{xy} , perchè talvolta non ha significato fisico mentre in altri casi con un opportuno cambiamento di variabile essa può diventare nulla.

1.1.1 Operatori Differenziali

Spesso le equazioni alle derivate parziali sono rappresentate in forma più compatta utilizzando determinati operatori differenziali, tra i quali:

1. Il **Gradiente** di $u(x, y, t)$:

$$\operatorname{gradu}(x, y, t) = \nabla u(x, y, t) = \begin{pmatrix} u_x(x, y, t) \\ u_y(x, y, t) \\ u_t(x, y, t) \end{pmatrix}$$

2. La **Divergenza** di una funzione vettoriale $u(x, y, t) = (u_1, u_2, u_3)$:

$$\operatorname{div}u(x, y, t) = \frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y} + \frac{\partial u_3}{\partial t}.$$

3. Il **Laplaciano** di $u(x, y, t)$:

$$\Delta u = \nabla^2 u(x, y, t) = \operatorname{div}(\operatorname{grad}(u(x, y, t))) = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial t^2}.$$

Descriviamo ora i più noti esempi di equazioni del secondo ordine.

Esempio 1.1.1 *L'equazione d'onda*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2}$$

descrive, in funzione della posizione e del tempo, lo spostamento, rispetto al punto di equilibrio, di una corda vibrante. L'equazione descrive anche il campo elettrico o magnetico in un'onda elettromagnetica oppure l'intensità di corrente oppure il potenziale lungo una linea di trasmissione. La quantità c è la velocità di propagazione dell'onda.

Esempio 1.1.2 *L'equazione di diffusione*

$$\kappa \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) = \frac{\partial u}{\partial t}$$

descrive la temperatura in una regione che non contiene sorgenti di calore, e si applica anche alla diffusione di un composto chimico in un mezzo permeabile (liquido, mezzo poroso) avente concentrazione $u(x, y, t)$. La costante κ viene detta diffusività.

Esempio 1.1.3 *L'equazione di Laplace*

$$\nabla^2 u = 0 \quad \Leftrightarrow \quad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

può essere ottenuta dall'equazione di diffusione ponendo a zero la derivata rispetto al tempo e descrive la distribuzione di temperatura in regime stazionario di un solido privo di sorgenti di calore. L'equazione di Laplace descrive anche il potenziale elettrostatico in una regione priva di carica elettrica. Si applica anche al flusso di un fluido incompressibile in una regione senza vortici, sorgenti o scarichi.

Esempio 1.1.4 *L'equazione di Poisson*

$$\nabla^2 u = \rho(x, y) \quad \Leftrightarrow \quad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \rho(x, y)$$

descrive la stessa situazione dell'equazione di Laplace ma in una regione in cui c'è carica elettrica oppure una sorgente di calore o di fluido. La funzione $\rho(x, y)$ si chiama densità di sorgente e dipende anche da costanti fisiche. Per esempio nell'equazione di Poisson:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = -\frac{\rho(x, y)}{\varepsilon}$$

dove $u(x, y)$ rappresenta il potenziale elettrostatico di una regione dello spazio, $\rho(x, y)$ rappresenta la densità della carica elettrica e ε è la permittività della sostanza.

Esempio 1.1.5 *Studiando il fenomeno della filtrazione dell'acqua nel sottosuolo supponiamo che Ω sia una regione dello spazio occupata da un cosiddetto mezzo poroso (terra o argilla per esempio), che abbia conduttività idraulica K . Indicando con $\phi(x, y, z)$ il livello dell'acqua, e con*

$$\mathbf{q} = (q_x, q_y, q_z)$$

la velocità di filtrazione, allora applicando la legge di Darcy risulta che tale velocità è proporzionale alla variazione del livello dell'acqua:

$$\mathbf{q} = -K \left(\frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial y}, \frac{\partial \phi}{\partial z} \right)$$

e inoltre, per la proprietà di conservazione della massa, la divergenza di \mathbf{q} deve essere nulla:

$$\operatorname{div} \mathbf{q} = 0.$$

Applicando la definizione di divergenza segue:

$$\begin{aligned} \operatorname{div} \mathbf{q} &= \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z} = \\ &= -K \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} \right) = 0. \end{aligned}$$

Da cui segue che la funzione ϕ soddisfa l'equazione di Laplace:

$$\Delta \phi = 0.$$

La motivazione che spinge a risolvere numericamente le equazioni alle derivate parziali sta nel fatto che non esistono tecniche analitiche generali per risolverle e anche se per alcuni tipi di equazioni (soprattutto lineari) si può scrivere l'espressione esplicita della soluzione teorica sotto forma di serie di Fourier. Tali serie convergono con grande lentezza quindi per ottenere una buona approssimazione della soluzione si richiede il calcolo di un numero di termini particolarmente elevato. Per altri tipi di equazioni si può scrivere l'espressione della soluzione teorica sotto forma di integrali che richiedono comunque un'approssimazione numerica.

Considerando un'equazione alle derivate parziali del secondo ordine essa può essere di tipo **Lineare**:

$$a(x, y)u_{xx} + 2b(x, y)u_{xy} + c(x, y)u_{yy} + d(x, y)u_x + e(x, y)u_y + f(x, y)u + g(x, y) = 0.$$

oppure **Quasi-Lineare**:

$$a(x, y, u, u_x, u_y)u_{xx} + 2b(x, y, u, u_x, u_y)u_{xy} + c(x, y, u, u_x, u_y)u_{yy} + f(x, y, u, u_x, u_y) = 0.$$

oppure ancora **Semi-Lineare**:

$$a(x, y)u_{xx} + 2b(x, y)u_{xy} + c(x, y)u_{yy} + f(x, y, u, u_x, u_y) = 0.$$

1.2 Classificazione delle equazioni alle derivate parziali

Consideriamo l'equazione alle derivate parziali lineare

$$au_{xx} + 2bu_{xy} + cu_{yy} + du_x + eu_y + fu + g = 0, \quad (1.1)$$

con $(x, y) \in \Omega \subset \mathbb{R}^2$ e tale che $a^2 + b^2 + c^2 \neq 0$ per ogni $(x, y) \in \Omega$. La classificazione delle equazioni alle derivate parziali avviene in base al segno assunto dalla quantità

$$\Delta = b^2 - ac.$$

Infatti un'equazione alle derivate parziali del secondo ordine si dice:

1. **iperbolica** se $\Delta > 0$,

2. **ellittica** se $\Delta < 0$,
3. **parabolica** se $\Delta = 0$.

Tale classificazione (per la verità un po' superata) dipende solo dall'analogia formale tra la (1.1) e l'equazione completa di una conica

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0,$$

che rappresenta

1. un' *iperbole* se $b^2 - ac > 0$,
2. un' *ellisse* se $b^2 - ac < 0$,
3. una *parabola* se $b^2 - ac = 0$.

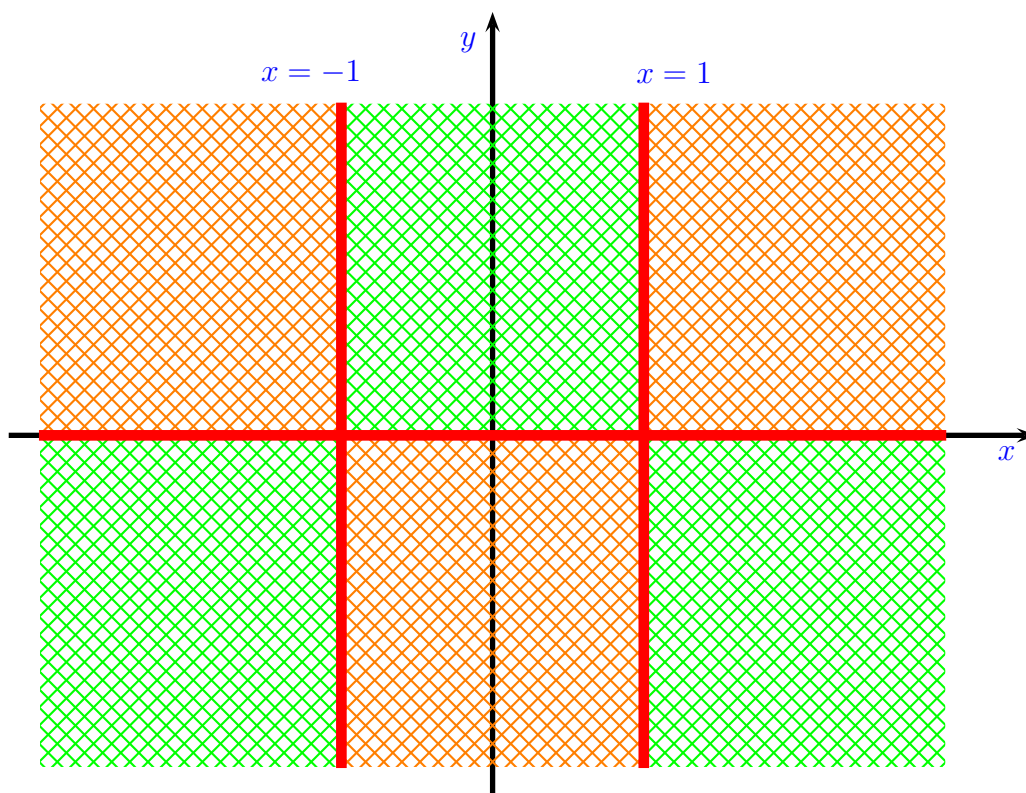
Si tratta di una classificazione non univoca perchè se i coefficienti a, b, c dipendono da x e y allora il tipo di equazione dipende dal dominio di definizione. Consideriamo per esempio l'equazione del secondo ordine:

$$y(x^2 + 1)u_{xx} + (x^2 - 1)u_{yy} + 3x + y = 0$$

con $(x, y) \in \Omega \subset \mathbb{R}^2$. In questo caso

$$\Delta = -y(x^2 - 1)(x^2 + 1)$$

e la situazione è schematizzata nel seguente grafico, in cui lungo le rette **rosse** l'equazione è parabolica, nella zona **verde** è iperbolica, mentre nella zona **arancio** ellittica.



Un modo indubbiamente migliore, e sicuramente univoco, per classificare un'equazione alle derivate parziali è quello di farlo in base al tipo di fenomeno che essa descrive. Le equazioni alle derivate parziali possono essere così divise in due tipi: *equazioni stazionarie*, in cui tutte le variabili sono spaziali, ed *equazioni di evoluzione*, le quali presentano una derivazione sia rispetto allo spazio che rispetto al tempo. Le equazioni di evoluzione rappresentano modelli che subiscono cambiamenti nel tempo e sono molto importanti nella descrizione dei fenomeni d'onda, dei fenomeni termodinamici, nei processi di diffusione e nella dinamica delle popolazioni. Le equazioni alle derivate parziali, come detto in precedenza, vengono classificate in ellittiche, paraboliche ed iperboliche. Le equazioni ellittiche sono di tipo stazionario mentre le paraboliche e le iperboliche sono equazioni di evoluzione. Le equazioni di evoluzione possono essere viste come delle equazioni differenziali ordinarie senza variabili spaziali, infatti uno dei metodi più efficaci per risolvere le equazioni di evoluzione è quello di approssimarle con un sistema di equazioni differenziali ordinarie.

Capitolo 2

Equazioni ellittiche

2.1 L'equazione di Laplace

Vediamo ora di descrivere una tecnica per la risoluzione numerica della più semplice equazione ellittica lineare che prende il nome di *equazione di Laplace*:

$$u_{xx} + u_{yy} = 0, \quad (2.1)$$

o, scritta alternativamente,

$$\Delta u \equiv \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

Se una funzione $u(x, y)$ è di classe \mathcal{C}^2 in un determinato sottoinsieme Ω di \mathbb{R}^2 ed è una soluzione di (2.1) nello stesso Ω allora prende il nome di *funzione armonica*. Le proprietà di queste funzioni sono state studiate con molta attenzione a causa della loro importanza nella teoria del potenziale e in quella gravitazionale. La principale tra queste è ununciata nel seguente teorema.

Teorema 2.1.1 (Principio del massimo) *Sia Ω una regione limitata e semplicemente connessa e sia Γ la sua frontiera. Sia $\bar{\Omega} = \Omega \cup \Gamma$. Se $u(x, y)$ è armonica su Ω e continua su $\bar{\Omega}$, allora $u(x, y)$ assume il suo valore massimo su Γ .*

L'equazione di Laplace può essere associata ad un problema di Dirichlet quando, assegnata una funzione $f(x, y)$ di classe $\mathcal{C}^2(\Gamma)$ si cerca una funzione $u(x, y)$ tale che:

1. $u(x, y)$ è continua su $\Omega \cup \Gamma$;
2. $u(x, y) = f(x, y)$ per ogni $(x, y) \in \Gamma$;
3. $u(x, y)$ è armonica in Ω .

In alternativa si può imporre la cosiddetta condizione di Neumann in cui, al posto della condizione 2., si impone che sia

$$\frac{\partial u}{\partial n} = f(x, y)$$

cioè sia assegnata la derivata normale di $u(x, y)$ rispetto alla curva Γ . Ricordiamo che se $\mathbf{n}^T = (n_x, n_y)$, è il vettore normale allora

$$\frac{\partial u}{\partial n} = n_x \frac{\partial u}{\partial x} + n_y \frac{\partial u}{\partial y}.$$

Consideriamo ora la risoluzione dell'equazione di Laplace prendendo Ω uguale al rettangolo $[a, b] \times [c, d]$, con $b > a$ e $d > c$. In questo caso un metodo è quello di approssimare l'operatore differenziale dopo avere suddiviso in modo opportuno l'insieme Ω . Infatti si suddivide l'intervallo $[a, b]$ in N parti uguali sull'asse x e M sull'asse y ottenendo la reticolazione di Ω , ponendo $x_0 = a$ e $y_0 = c$ e definendo i seguenti punti:

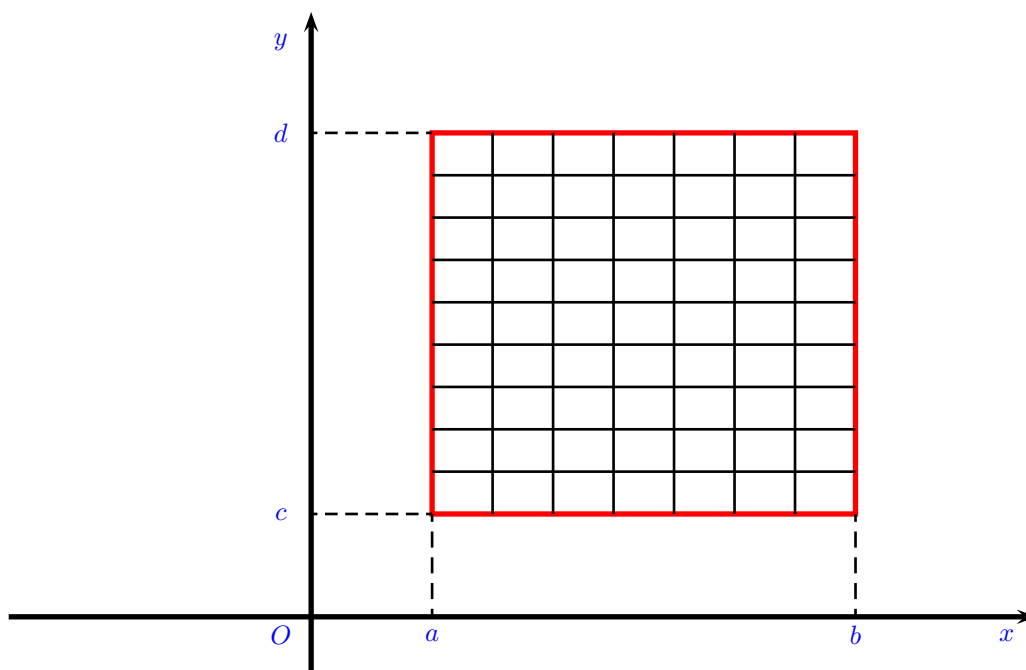
$$x_i = x_{i-1} + k = a + ik \quad i = 1, 2, \dots, N$$

$$y_j = y_{j-1} + h = c + jh \quad j = 1, 2, \dots, M$$

dove $h = (b - a)/N$ e $k = (d - c)/M$. Abbiamo così ottenuto un insieme discreto di punti del piano

$$\mathcal{R}_{N+1, M+1} = \{(x_i, y_j) \in \mathbb{R}^2 | x_i = a + ik, i = 0, N, y_j = c + jh, j = 0, M\}.$$

La risoluzione numerica del problema di Dirichlet associato consiste nell'approssimare opportunamente la funzione $u(x, y)$ nei punti appartenenti all'insieme $\mathcal{R}_{N+1, M+1}$, tenendo presente che la soluzione è nota sul perimetro del rettangolo $[a, b] \times [c, d]$.



2.2 Derivazione numerica

Prima di procedere nella descrizione del metodo vediamo come approssimare numericamente le derivate prima e seconda di una funzione $f(t)$, che supponiamo continua e differenziabile fino ad un opportuno ordine k . Suddividiamo l'intervallo di variabilità della t in sottointervalli di ampiezza h . Consideriamo tre punti consecutivi appartenenti a tale reticolazione, rispettivamente t_{i-1} , t_i e t_{i+1} tali che

$$t_{i-1} = t_i - h, \quad t_{i+1} = t_i + h.$$

Sviluppiamo la funzione $f(t_{i+1})$ in serie di Taylor prendendo come punto iniziale t_i :

$$f(t_{i+1}) = f(t_i) + hf'(t_i) + \frac{h^2}{2}f''(t_i) + \frac{h^3}{6}f'''(t_i) + \frac{h^4}{24}f^{iv}(\xi_i), \quad \xi_i \in [t_i, t_{i+1}]$$

e procediamo in modo analogo per $f(t_{i-1})$:

$$f(t_{i-1}) = f(t_i) - hf'(t_i) + \frac{h^2}{2}f''(t_i) - \frac{h^3}{6}f'''(t_i) + \frac{h^4}{24}f^{iv}(\eta_i), \quad \eta_i \in [t_{i-1}, t_i].$$

Sommiamo ora le due espressioni

$$f(t_{i+1}) + f(t_{i-1}) = 2f(t_i) + h^2 f''(t_i) + \frac{h^4}{24} [f^{iv}(\xi_i) + f^{iv}(\eta_i)]$$

ricavando

$$f''(t_i) = \frac{f(t_{i+1}) - 2f(t_i) + f(t_{i-1}))}{h^2} - \frac{h^2}{24} [f^{iv}(\xi_i) + f^{iv}(\eta_i)]$$

e, trascurando l'ultimo termine, l'approssimazione della derivata seconda è:

$$f''(t_i) \simeq \frac{f(t_{i+1}) - 2f(t_i) + f(t_{i-1}))}{h^2} \quad (2.2)$$

mentre si può provare che l'errore vale:

$$E(f''(t_i)) = -\frac{h^2}{12} f^{iv}(\xi), \quad \xi \in [t_{i-1}, t_{i+1}].$$

Poniamoci lo stesso problema per la derivata prima e procediamo nello stesso modo cioè scrivendo le serie di Taylor per $f(t_{i-1})$ e $f(t_{i+1})$:

$$f(t_{i+1}) = f(t_i) + hf'(t_i) + \frac{h^2}{2} f''(t_i) + \frac{h^3}{6} f'''(\sigma_i), \quad \sigma_i \in [t_i, t_{i+1}]$$

$$f(t_{i-1}) = f(t_i) - hf'(t_i) + \frac{h^2}{2} f''(t_i) - \frac{h^3}{6} f'''(\mu_i), \quad \mu_i \in [t_{i-1}, t_i]$$

e questa volta sottraiamo la seconda dalla prima:

$$f(t_{i+1}) - f(t_{i-1}) = 2hf'(t_i) + \frac{h^3}{6} [f'''(\sigma_i) + f'''(\mu_i)]$$

ottenendo

$$f'(t_i) = \frac{f(t_{i+1}) - f(t_{i-1}))}{2h} - \frac{h^2}{12} [f'''(\sigma_i) + f'''(\mu_i)]$$

e, trascurando l'ultimo termine, l'approssimazione della derivata prima è:

$$f'(t_i) \simeq \frac{f(t_{i+1}) - f(t_{i-1}))}{2h} \quad (2.3)$$

mentre si può provare che l'errore vale:

$$E(f'(t_i)) = -\frac{h^2}{6} f'''(\delta), \quad \delta \in [t_{i-1}, t_{i+1}].$$

La formula (2.3) prende il nome di *formula alle differenze centrali*. Osserviamo che sia per questa che per l'approssimazione numerica per la derivata seconda l'errore dipende da h^2 , sono formule cioè *del secondo ordine*. Vediamo ora altre due approssimazioni per la derivata prima. Infatti possiamo anche scrivere:

$$f(t_{i+1}) = f(t_i) + hf'(t_i) + \frac{h^2}{2}f''(\xi_i), \quad \xi_i \in [t_i, t_{i+1}]$$

da cui si ricava immediatamente la *formula alle differenze in avanti*:

$$f'(t_i) \simeq \frac{f(t_{i+1}) - f(t_i)}{h}$$

con errore

$$E(f'(t_i)) = -\frac{h}{2}f''(\xi_i),$$

e, analogamente,

$$f(t_{i-1}) = f(t_i) - hf'(t_i) + \frac{h^2}{2}f''(\mu_i), \quad \mu_i \in [t_{i-1}, t_i]$$

da cui si ricava immediatamente la *formula alle differenze all'indietro*:

$$f'(t_i) \simeq \frac{f(t_i) - f(t_{i-1})}{h}$$

con errore

$$E(f'(t_i)) = -\frac{h}{2}f''(\mu_i).$$

Queste due formule hanno ordine 1, quindi sono meno precise rispetto alla formula alle differenze centrali, tuttavia hanno il pregio di poter essere applicate quando la derivata prima è discontinua in t_i . Torniamo ora alla risoluzione numerica dell'equazione di Laplace. L'idea alla base del metodo è quella di approssimare le derivate parziali seconde nei punti del reticolo $\mathcal{R}_{N+1, M+1}$ e imporre che tali approssimazioni soddisfino l'equazione di Laplace. Poniamo innanzitutto

$$u_{i,j} \simeq u(x_i, y_j), \quad i = 0, 1, \dots, N, \quad j = 0, 1, \dots, M$$

e, per approssimare la derivata parziale seconda $u_{xx}(x_i, y_j)$, consideriamo i seguenti 3 punti del reticolo (x_{i-1}, y_j) , (x_i, y_j) e (x_{i+1}, y_j) e, applicando la formula (2.2) supponendo costante il valore y_j , risulta:

$$\frac{\partial^2 u}{\partial x^2}(x_i, y_j) \simeq \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2}.$$

Analogamente per approssimare $u_{yy}(x_i, y_j)$ consideriamo i seguenti 3 punti del reticolo (x_i, y_{j-1}) , (x_i, y_j) e (x_i, y_{j+1}) e, applicando la formula (2.2), risulta:

$$\frac{\partial^2 u}{\partial y^2}(x_i, y_j) \simeq \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{k^2}.$$

Tenendo presente che la funzione $u(x, y)$ è nota sul bordo del rettangolo alcune delle approssimazioni non devono essere calcolate, infatti:

$$\begin{aligned} u_{0,j} &= u(a, y_j) = f(a, y_j), & j &= 0, \dots, M \\ u_{i,0} &= u(x_i, c) = f(x_i, c), & i &= 0, \dots, N \\ u_{N,j} &= u(x_N, y_j) = f(b, y_j), & j &= 0, \dots, M \\ u_{i,M} &= u(x_i, y_M) = f(x_i, d), & i &= 0, \dots, N. \end{aligned}$$

Adesso possiamo imporre che queste approssimazioni soddisfano l'equazione di Laplace

$$\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{k^2} = 0$$

$$(u_{i+1,j} - 2u_{i,j} + u_{i-1,j})k^2 + (u_{i,j+1} - 2u_{i,j} + u_{i,j-1})h^2 = 0$$

Il metodo numerico, che viene detto anche **Metodo a cinque punti** assume quindi la forma finale

$$h^2 u_{i,j-1} + k^2 u_{i-1,j} - 2(h^2 + k^2)u_{i,j} + k^2 u_{i+1,j} + h^2 u_{i,j+1} = 0$$

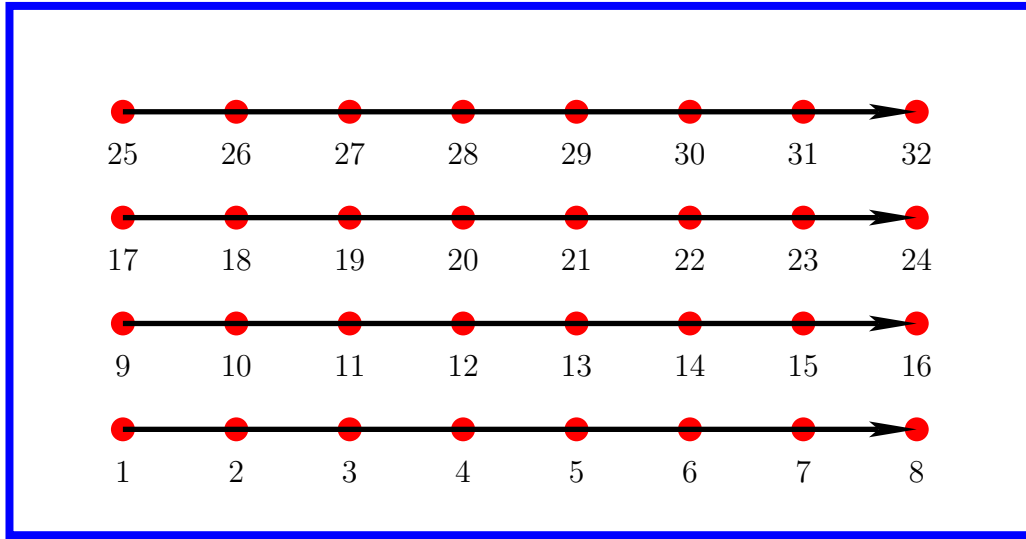
con $i = 1, \dots, N - 1$, e $j = 1, \dots, M - 1$.

2.2.1 Ordinamento delle incognite

Le relazioni che legano le incognite $u_{i,j}$ formano un sistema lineare la cui forma dipende dal modo con cui vengono ordinate tali incognite. Un primo modo di riordinare le incognite $u_{i,j}$ è quello di porre

$$\mathbf{u}^T = (u_{1,1}, u_{2,1}, \dots, u_{N-1,1}, u_{1,2}, \dots, u_{N-1,2}, \dots, u_{1,M-1}, \dots, u_{N-1,M-1})$$

ed è schematizzato dal seguente grafico.



Tale ordinamento prende il nome di **Ordinamento naturale (o lessicografico)**. Si inizia quindi dal punto (x_1, y_1) e si procede verso destra. Appena terminata la riga si passa a quella superiore, si considera il punto (x_1, y_2) e così via. La prima equazione si ottiene per $i = j = 1$:

$$k^2 u_{1,0} + h^2 u_{0,1} - 2(h^2 + k^2)u_{1,1} + h^2 u_{2,1} + k^2 u_{1,2} = 0$$

equivalente a

$$-2(h^2 + k^2)u_{1,1} + h^2 u_{2,1} + k^2 u_{1,2} = -k^2 u_{1,0} - h^2 u_{0,1}.$$

La seconda equazione si ottiene per $i = 2$ e $j = 1$:

$$k^2 u_{2,0} + h^2 u_{1,1} - 2(h^2 + k^2)u_{2,1} + h^2 u_{3,1} + k^2 u_{2,2} = 0$$

equivalente a

$$h^2 u_{1,1} - 2(h^2 + k^2)u_{2,1} + h^2 u_{3,1} + k^2 u_{2,2} = -k^2 u_{2,0}.$$

Ogni equazione (i, j) ha al più 5 coefficienti diversi da 0 di cui 3 coinvolgono 3 incognite numerata consecutivamente $(i - 1, j)$, (i, j) e $(i + 1, j)$, una precedente $(i, j - 1)$ e una successiva $(i, j + 1)$, distanti $N - 1$ incognite (prima e dopo quella di riferimento). Per ottenere le approssimazioni è quindi necessario risolvere un sistema lineare

$$A\mathbf{u} = \mathbf{b}$$

che ha la seguente struttura tridiagonale a blocchi

$$A = \begin{pmatrix} T & J & & & \\ J & T & J & & \\ & \ddots & \ddots & \ddots & \\ & & & J & T & J \\ & & & & J & T \end{pmatrix}$$

dove $J = k^2 I_{N-1}$, essendo I_{N-1} è la matrice identità di ordine $N - 1$, e T è la seguente matrice tridiagonale di dimensione $N - 1$:

$$T = \begin{pmatrix} -2(h^2 + k^2) & h^2 & & & & \\ h^2 & -2(h^2 + k^2) & h^2 & & & \\ & \ddots & \ddots & \ddots & & \\ & & h^2 & -2(h^2 + k^2) & h^2 & \\ & & & h^2 & -2(h^2 + k^2) \end{pmatrix}.$$

Nella Figura 2.1 è riportata la struttura della matrice nel caso della griglia riportata come esempio in precedenza. Il sistema lineare da risolvere ha una struttura molto sparsa: se indichiamo con n la sua dimensione ($n = (M - 1)(N - 1)$) poco meno di $5n$ elementi sono diversi da zero su n^2 elementi della matrice dei coefficienti. Questo significa che il sistema può essere risolto in modo più efficiente utilizzando i cosiddetti metodi iterativi (vedere Capitolo 5) e non diretti (tipo fattorizzazione LU) che distruggerebbero la struttura sparsa della matrice. Come esempio di questa affermazione nella Figura 2.2 riportiamo la struttura della matrice triangolare superiore U della fattorizzazione LU di A , come si può osservare ha ben 231 elementi diversi da zero (da confrontare con i 136 di A).

Altri ordinamenti delle incognite sono:

1. **Ordinamento Cuthill-McKee;**
2. **Ordinamento Red-Black;**
3. **Ordinamento Multicolore.**

Nell'ordinamento Cuthill-McKee, proposto nel 1969, si ordinano le incognite partendo da un punto fissato e numerando i punti adiacenti privilegiando quelli che si trovano lungo una determinata direzione (per esempio quella diagonale) se non ce ne sono si prende un altro nodo adiacente e si prosegue. La tecnica è schematizzata nel seguente grafico.

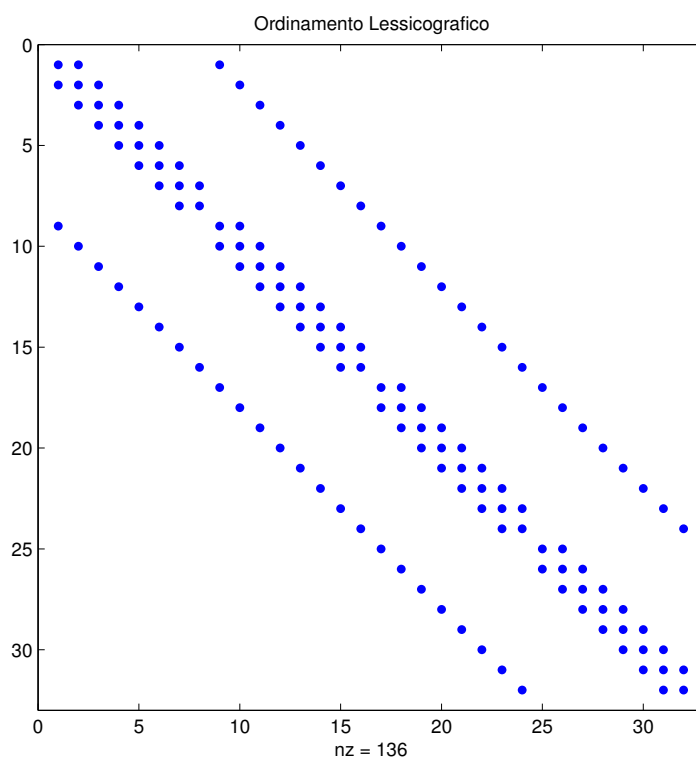
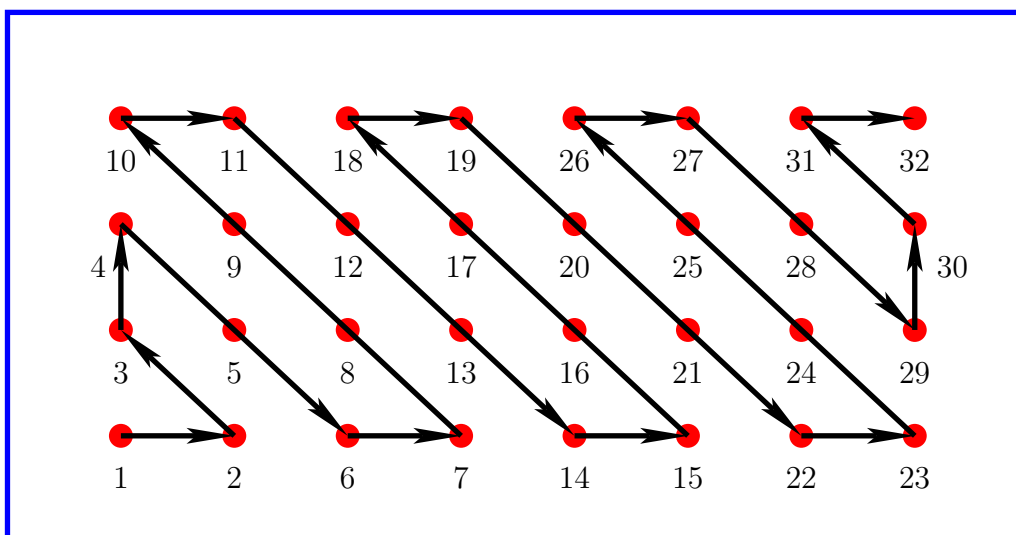


Figura 2.1: Struttura della matrice dei coefficienti per l'ordinamento Lessicografico



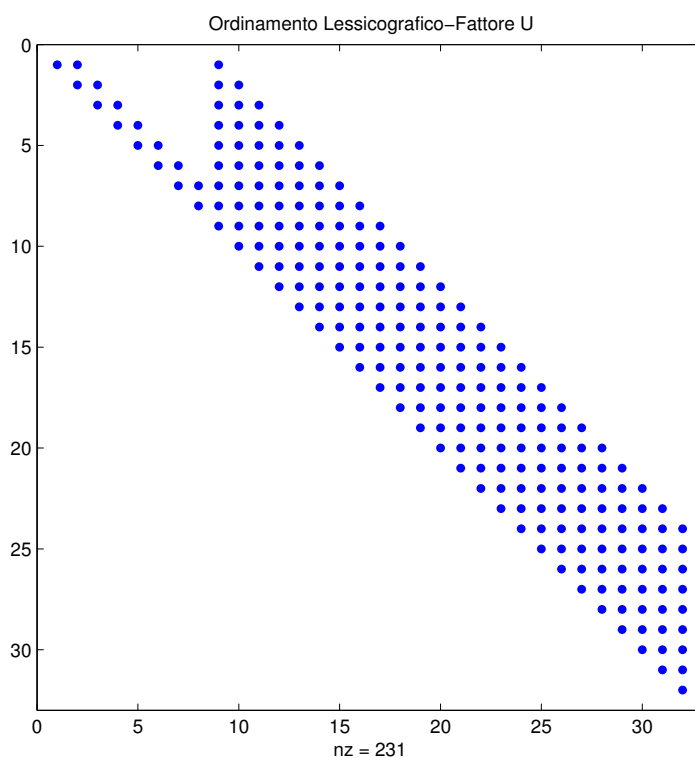


Figura 2.2: Struttura del fattore triangolare U per la matrice dei coefficienti definita dall'ordinamento Lessicografico

L'obiettivo è quello di diminuire l'ampiezza di banda della matrice. Infatti in Figura 2.3 è visualizzata la struttura della matrice A con tale ordinamento mentre la Figura 2.4 mostra la struttura della matrice U triangolare superiore. Osserviamo come il numero di elementi diversi da zero sia piuttosto basso (148) rispetto al caso precedente.

Altri due ordinamenti, cui accenniamo per motivi di completezza prevedono che i nodi vengano suddivisi (colorati) in due (o più) tipi. Nell'ordinamento Red-Black sono divisi in rossi e neri, in modo tale che due nodi adiacenti abbiano colori diversi. Finita la colorazione i nodi sono numerati usando l'ordinamento lessicografico applicato ad un colore per volta, come si evidenzia nel seguente schema.

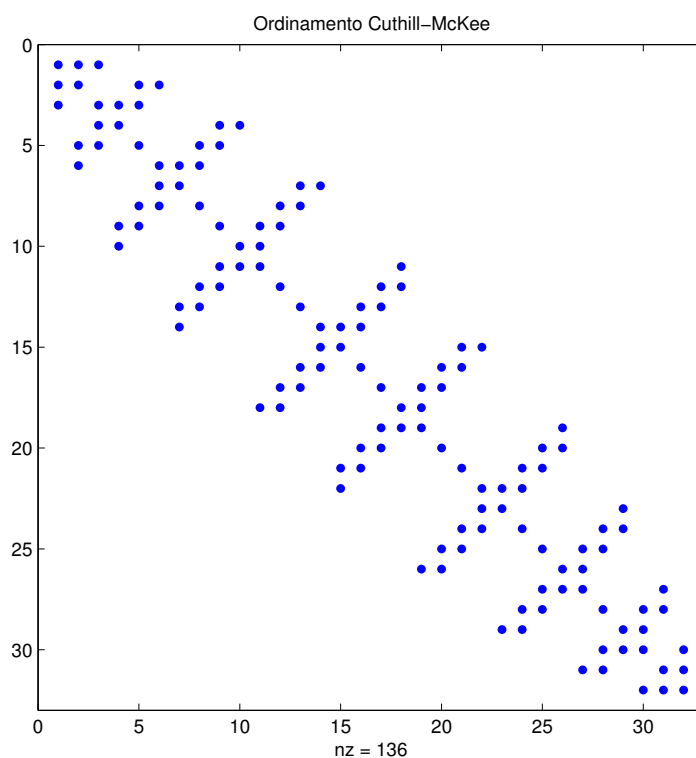
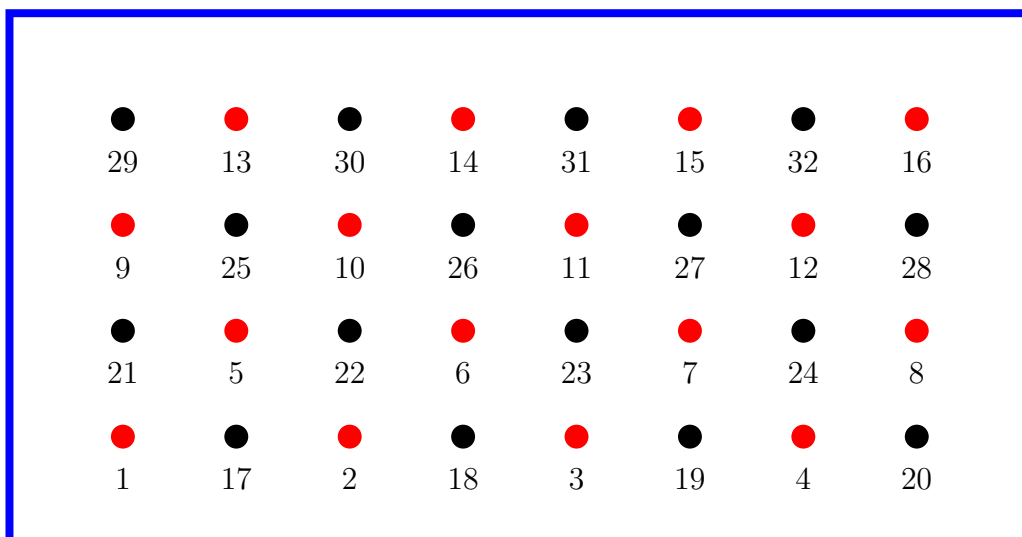


Figura 2.3: Struttura della matrice dei coefficienti per l'ordinamento Cuthill-McKee



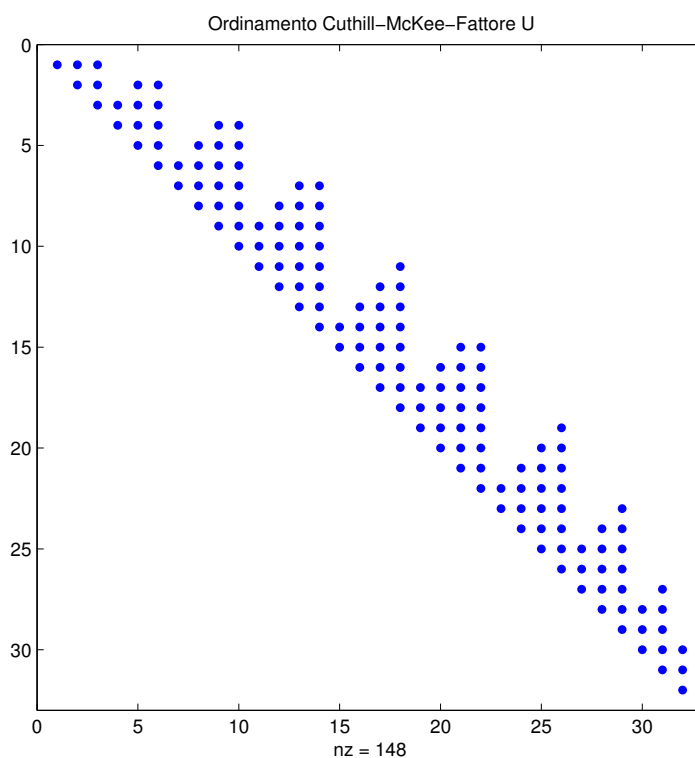


Figura 2.4: Struttura del fattore triangolare U per la matrice dei coefficienti definita dall'ordinamento Cuthill-McKee

In Figura 2.5 è visualizzata la struttura della matrice A con tale ordinamento mentre la Figura 2.6 mostra la struttura della matrice U triangolare superiore. Osserviamo come il numero di elementi diversi da zero (170) sia più alto rispetto all'ordinamento Cuthill-McKee ma inferiore rispetto all'ordinamento lessicografico.

L'ordinamento multicolore è uguale al Red-Black ma si usano più colori, solitamente 4 o 6, e comunque un numero pari. I nodi sono ordinati colorandoli in sequenza uno di ogni colore diverso, mentre per la riga successiva si parte dalla coppia di colori successiva, in modo tale che su ogni colonna ci siano solo due colori. Nella figura seguente vediamo un esempio con 4 colori.

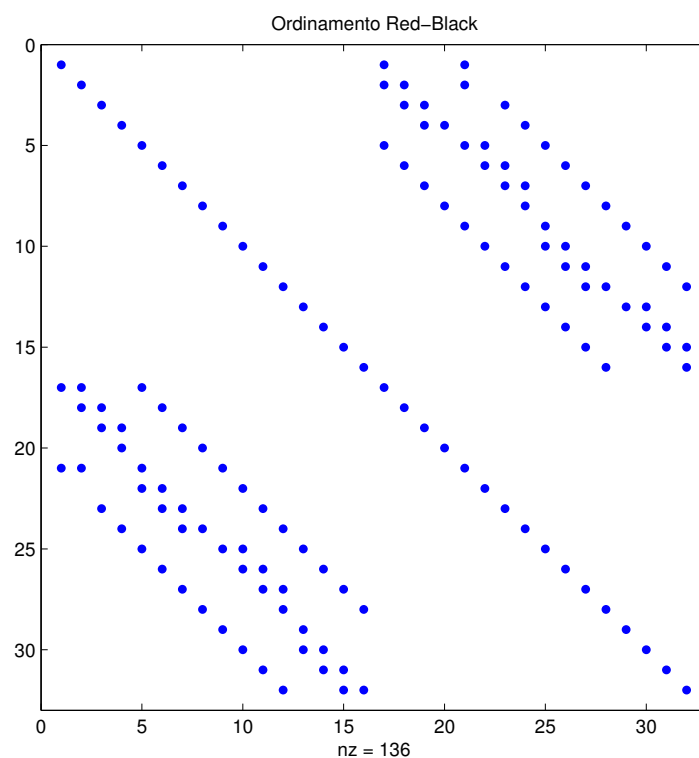
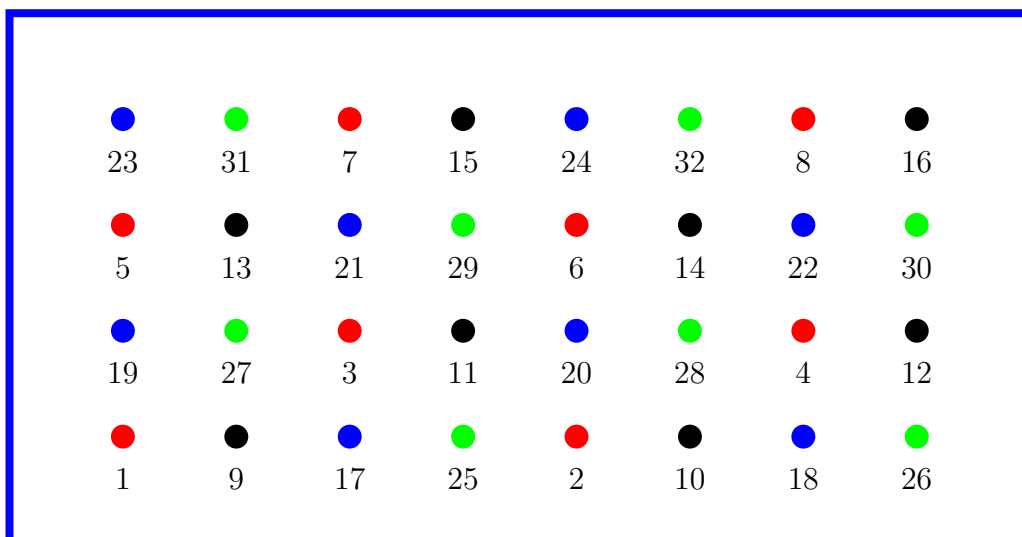


Figura 2.5: Struttura della matrice dei coefficienti per l'ordinamento Red-Black



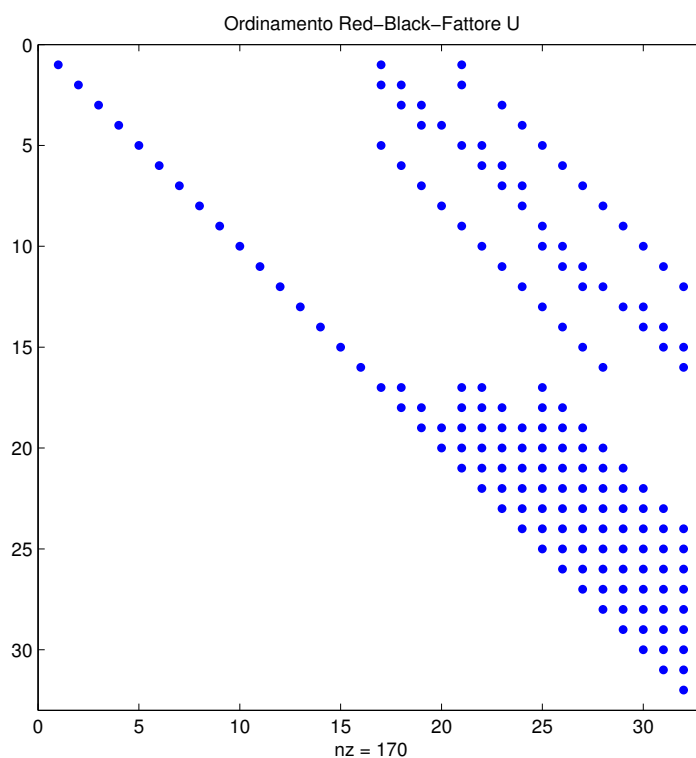


Figura 2.6: Struttura del fattore triangolare U per la matrice dei coefficienti definita dall'ordinamento Red-Black

In Figura 2.7 è visualizzata la struttura della matrice A con tale ordinamento mentre la Figura 2.8 mostra la struttura della matrice U triangolare superiore. Osserviamo come il numero di elementi diversi da zero (191) sia più alto rispetto agli ordinamenti Cuthill-McKee e Red-Black ma inferiore rispetto all'ordinamento lessicografico.

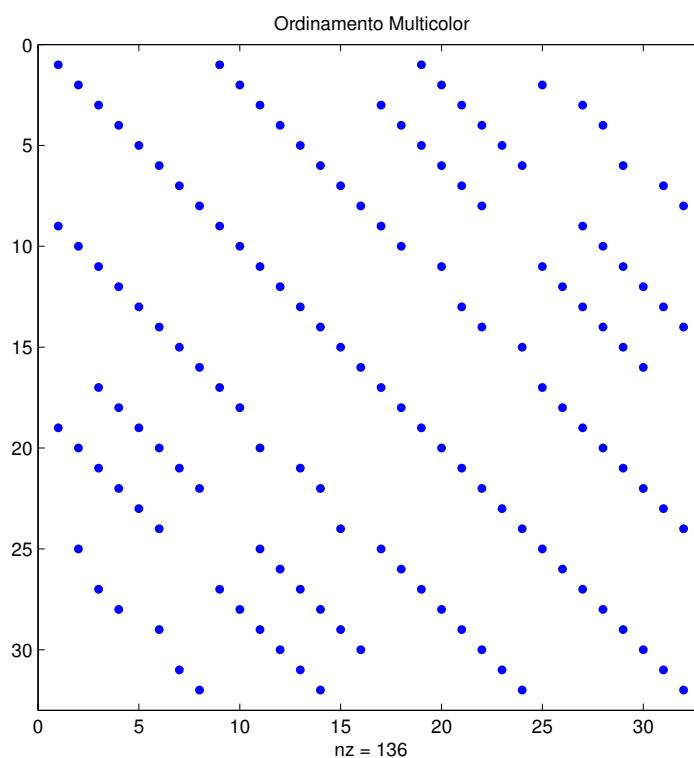


Figura 2.7: Struttura della matrice dei coefficienti per l'ordinamento a 4 colori

2.3 Equazione di Laplace su domini irregolari

L'uso delle differenze divise funziona bene quando il dominio Ω è un rettangolo, oppure un quadrato o un poligono che può essere scomposto come un'unione di quadrati o rettangoli. Quando invece il contorno Γ del dominio di integrazione Ω è un poligono oppure una curva regolare a tratti l'approssimazione delle derivate parziali è piuttosto problematico. C'è solo un caso in cui è possibile ricondursi ad un dominio rettangolare ed è quello in cui Γ è una circonferenza, per esempio

$$\Gamma = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}.$$

e Ω è il cerchio

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$$

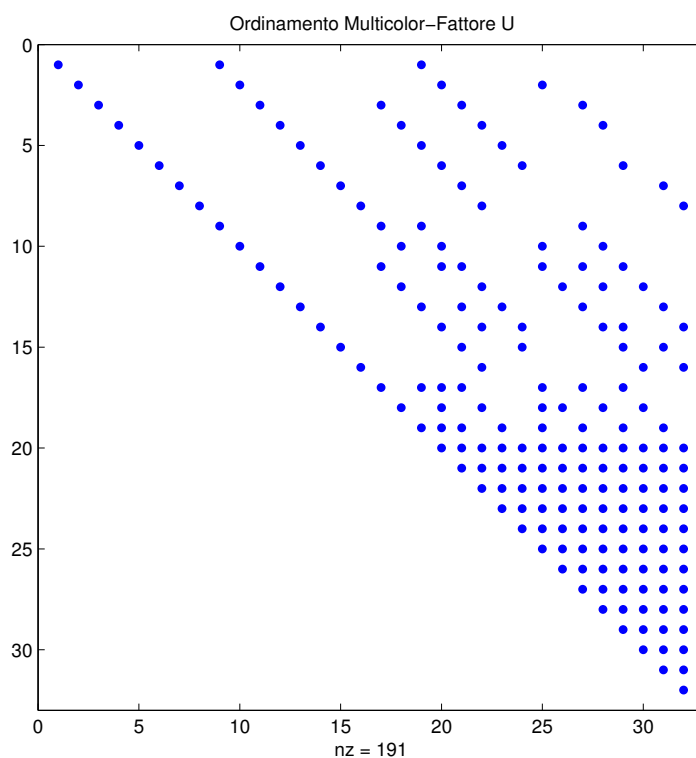


Figura 2.8: Struttura del fattore triangolare U per la matrice dei coefficienti definita dall'ordinamento Multicolore con 4 colori

In questo caso il dominio può essere trasformato in un rettangolo cambiando le coordinate cartesiane in polari:

$$x = r \cos \theta, \quad y = r \sin \theta.$$

In questo caso il dominio diventa

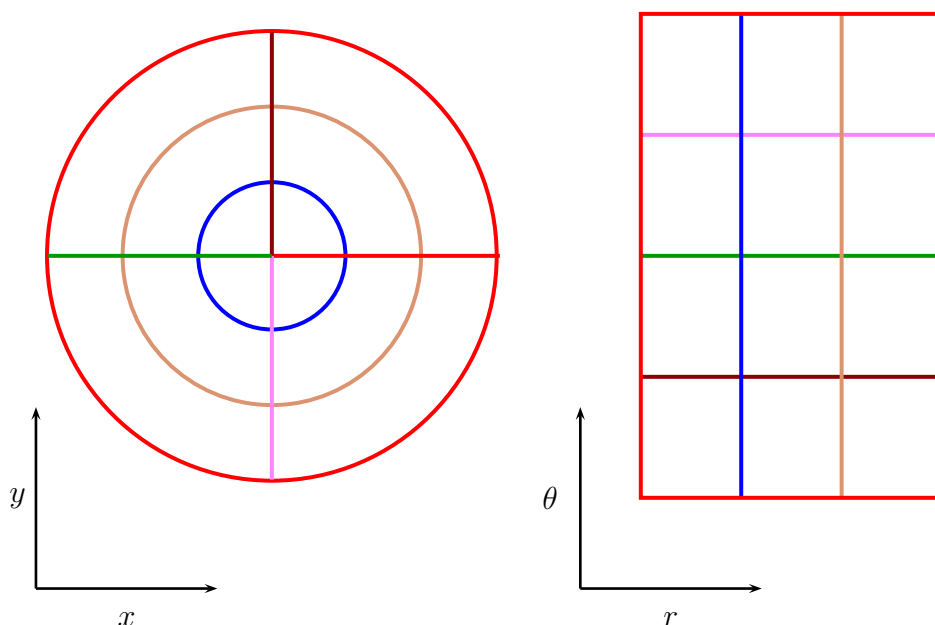
$$\Omega = \{(r, \theta) \in \mathbb{R}^2 : 0 \leq r < 1, 0 \leq \theta < 2\pi\}.$$

e

$$\Gamma = \{(r, \theta) \in \mathbb{R}^2 : r = 1\}.$$

Per evitare confusione poniamo

$$v(r, \theta) = u(r \cos \theta, r \sin \theta).$$



La condizione al contorno è assegnata ovviamente su Γ cioè:

$$u(x, y) = u(\cos \theta, \sin \theta) = f(\cos \theta, \sin \theta).$$

L'equazione di Laplace in coordinate polari diventa

$$\frac{\partial^2 v}{\partial r^2} + \frac{1}{r^2} \frac{\partial^2 v}{\partial \theta^2} + \frac{1}{r} \frac{\partial v}{\partial r} = 0.$$

Il problema più delicato riguarda la trasformazione delle condizioni iniziali in coordinate polari. Il caso più semplice è il segmento $r = 1$ che coincide con la circonferenza Γ in coordinate cartesiane, quindi

$$v(1, \theta) = f(\cos \theta, \sin \theta), \quad 0 \leq \theta \leq 2\pi.$$

I segmenti che si ottengono per $0 < r < 1$ e $\theta = 0$ e $0 < r < 1$ e $\theta = 2\pi$ sono esattamente lo stesso segmento nel cerchio originale. Il valore della funzione su tale segmento non è noto perciò è necessario assegnare una cosiddetta *condizione di periodicità*:

$$v(r, 0) = v(r, 2\pi), \quad 0 < r < 1.$$

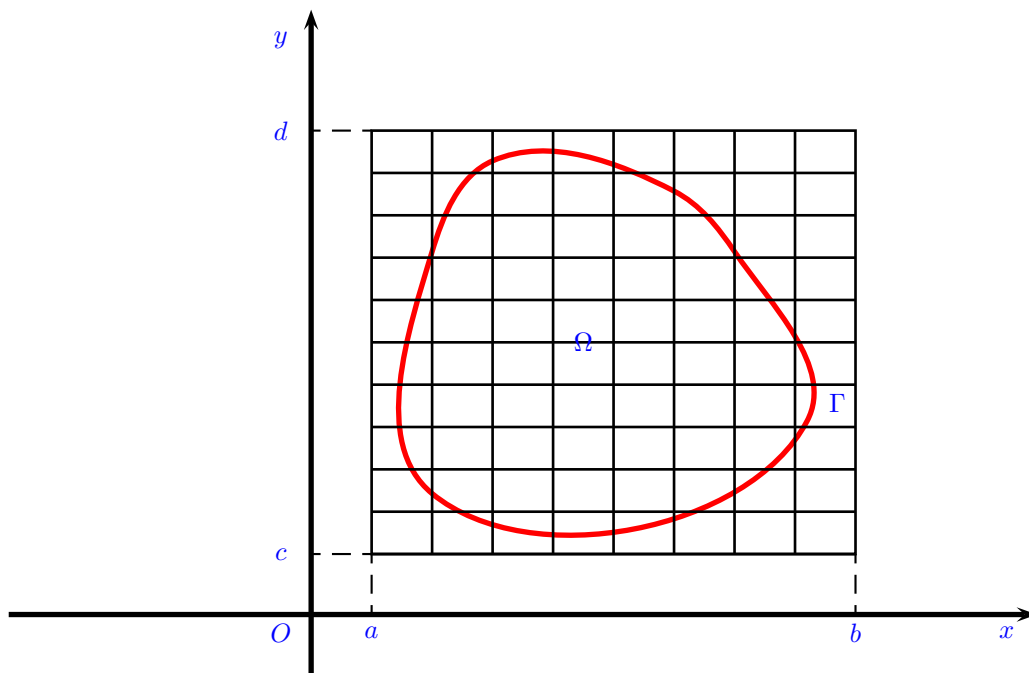
L'ultimo segmento da considerare è quello che si ottiene assegnando il valore $r = 0$, $0 \leq \theta \leq 2\pi$. L'intero segmento corrisponde ad un unico punto, cioè l'origine del piano cartesiano. Quindi la funzione v deve essere costante lungo tale linea e, per esprimere tale condizione in modo matematicamente più comprensibile poniamo la condizione di Neumann:

$$\frac{\partial v}{\partial \theta}(0, \theta) = 0.$$

Adesso si può risolvere numericamente l'equazione di Laplace in coordinate polari usando le approssimazioni alle differenze divise per le derivate parziali ed imponendo le condizioni al contorno con grande cautela.

Analizziamo ora il caso in cui la frontiera del dominio Ω sia una curva chiusa e regolare senza una particolare forma. In questo caso si considera un rettangolo $[a, b] \times [c, d]$ tale da contenere sia Ω che Γ e si discretizza tale regione come già visto in precedenza, definendo l'insieme discreto:

$$\mathcal{R}_{N+1, M+1} = \{(x_i, y_j) \in [a, b] \times [c, d] | i = 0, \dots, N, j = 0, \dots, M\}.$$



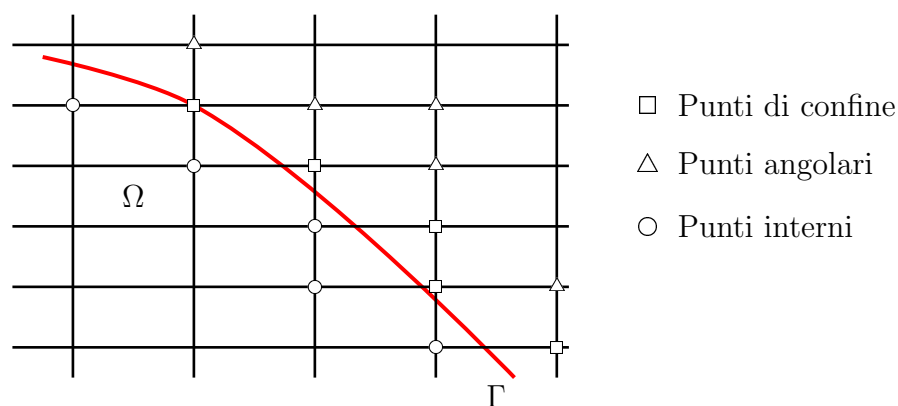
L'insieme dei punti discreti che appartengono sia al rettangolo $[a, b] \times [c, d]$ che al dominio Ω si denota con

$$\Omega_{hk} = \mathcal{R}_{N+1, M+1} \cap \Omega.$$

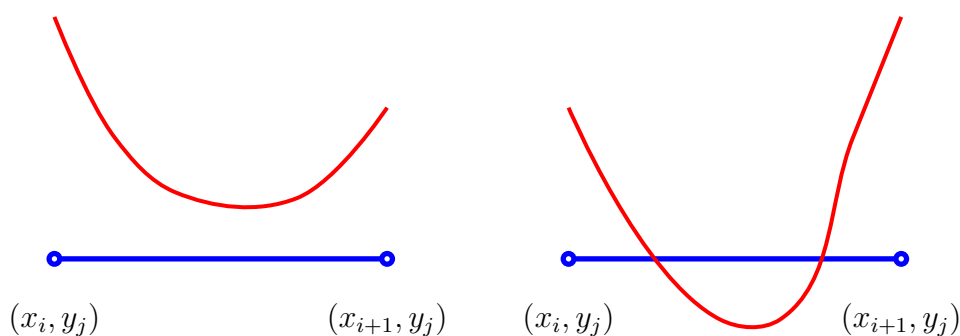
Tale insieme viene detto *insieme dei punti interni*. Ogni punto interno (x_i, y_j) ha quattro punti vicini, cioè $(x_{i\pm 1}, y_j)$ e $(x_i, y_{j\pm 1})$.

Un punto vicino ad un punto interno che non appartiene a Ω_{hk} si dice *punto di confine*. L'insieme dei punti di confine si indica con Γ_{hk} .

I *punti angolari* sono invece i punti dell'insieme $\mathcal{R}_{N+1, M+1}$ che non sono nè interni nè di confine, ma risultano essere vertici di una cella che contiene almeno un punto interno.



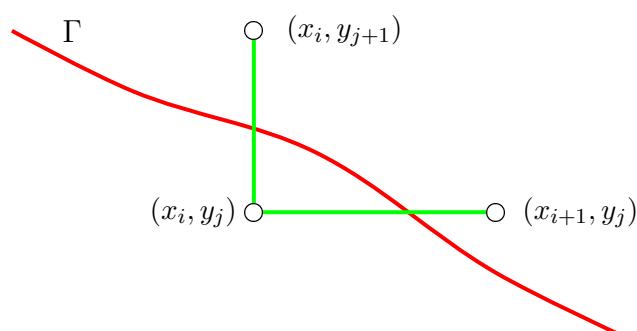
I punti angolari non hanno alcun ruolo nella soluzione numerica di equazioni ellittiche. L'ipotesi che si deve fare sulla griglia che si considera è che i segmenti che congiungono punti interni devono essere interamente contenuti nel dominio Ω . Si evita cioè il caso evidenziato dal secondo grafico:



Tale situazione può essere evitata scegliendo opportunamente il passo di discretizzazione oppure effettuando un opportuno cambio di variabile (per esempio una rotazione degli assi). Il problema dei domini irregolari sorge quando nelle approssimazioni delle derivate seconde intervengono valori della funzione calcolati nei punti di confine. Per ovviare a tale inconveniente ci sono diverse tecniche, due sono le più usate:

1. come valori nei punti di confine si utilizzano gli stessi valori assunti dalla condizione al contorno;
2. si utilizza il valore della condizione al contorno nel punto sulla curva Γ più vicino al punto interno.

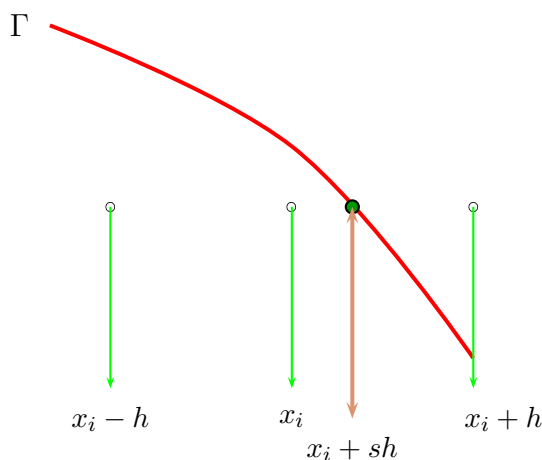
Nel primo caso, considerando la seguente situazione:



si pone:

$$u(x_i, y_{j+1}) = f(x_i, y_{j+1}), \quad u(x_{i+1}, y_j) = f(x_{i+1}, y_j).$$

Tale assegnazione rappresenta, dal punto di vista matematico, una forzatura, poichè in realtà non è noto neanche se la funzione $f(x, y)$ sia calcolabile in tali punti. Nel secondo caso, evidenziato dal seguente grafico, si utilizza il punto nel quale si conosce il valore della soluzione.



Approssimazione della derivata seconda su griglie non equidistanti

Consideriamo, a puro titolo di esempio, il problema di approssimare la derivata seconda della funzione $f(x)$ nel punto di ascissa x_i ma considerando i valori della funzione nei punti non equidistanti $x_i - h$ e $x_i + sh$, con $0 < s < 1$. Sviluppiamo in serie di Taylor la funzione nel punto $x_i + sh$ prendendo come punto iniziale x_i

$$f(x_i + sh) = f(x_i) + shf'(x_i) + \frac{s^2h^2}{2}f''(x_i) + \frac{s^3h^3}{6}f'''(\xi_i), \quad \xi_i \in [x_i, x_i + sh]$$

e procediamo in modo analogo per $f(x_{i-1})$:

$$f(x_{i-1}) = f(x_i) - hf'(x_i) + \frac{h^2}{2}f''(x_i) - \frac{h^3}{6}f'''(\eta_i), \quad \eta_i \in [x_{i-1}, x_i].$$

Moltiplichiamo per s quest'ultima espressione

$$sf(x_{i-1}) = sf(x_i) - shf'(x_i) + \frac{sh^2}{2}f''(x_i) - \frac{sh^3}{6}f'''(\eta_i)$$

e sommiamola con quella di $f(x_i + sh)$:

$$f(x_i + sh) + sf(x_{i-1}) = f(x_i)(1+s) + \frac{h^2}{2}f''(x_i)s(1+s) + \frac{sh^3}{6} [s^2f'''(\xi_i) - f'''(\eta_i)]$$

ricavando

$$f''(x_i) = 2 \frac{f(x_i + sh) - f(x_i)(1+s) + sf(x_{i-1})}{sh^2(1+s)} + \frac{h}{3(1+s)} [f'''(\eta_i) - s^2f'''(\xi_i)]$$

e, trascurando l'ultimo termine, l'approssimazione della derivata seconda è:

$$f''(x_i) \simeq 2 \frac{f(x_i + sh) - f(x_i)(1 + s) + sf(x_{i-1})}{sh^2(1 + s)} \quad (2.4)$$

mentre l'errore vale:

$$E(f''(x_i)) = \frac{h}{3(1 + s)} [f'''(\eta_i) - s^2 f'''(\xi_i)].$$

Applicando tale risultato ad una funzione in due variabili si otterrebbe l'approssimazione:

$$u_{xx}(x_i + sh, y_j) \simeq 2 \frac{u(x_i + sh, y_j) - (1 + s)u(x_i, y_j) + su(x_{i-1}, y_j)}{sh^2(1 + s)}.$$

Tale approssimazione risulta essere del primo ordine quindi meno precisa rispetto all'approssimazione della derivata seconda su punti equidistanti. Questo risultato ha come conseguenza un'approssimazione numerica meno precisa in prossimità del contorno del dominio di integrazione.

Volendo ottenere un'approssimazione più accurata si può utilizzare la seguente formula, della quale omettiamo la dimostrazione, che utilizza l'ulteriore valore $f(x_{i-2})$ ma è di ordine 2:

$$f''(x_i) \simeq \frac{1}{h^2} \left[\frac{s-1}{s+2} f(x_{i-2}) + \frac{2(2-s)}{s+1} f(x_{i-1}) - \frac{3-s}{s} f(x_i) + \frac{6f(x_i + sh)}{s(s+1)(s+2)} \right].$$

2.4 La generica equazione ellittica lineare

Una generica equazione ellittica lineare assume la forma

$$Au_{xx} + Cu_{yy} + P(x, y)u_x + Q(x, y)u_y + R(x, y)u + S(x, y) = 0 \quad (2.5)$$

con $AC > 0$. Con un opportuno cambio di scala si potrebbe porre $A = C = 1$ e ottenere così l'equazione

$$u_{xx} + u_{yy} + P(x, y)u_x + Q(x, y)u_y + R(x, y)u + S(x, y) = 0 \quad (2.6)$$

con $(x, y) \in \Omega$ e con condizione al contorno

$$u(x, y) = f(x, y), \quad (x, y) \in \Gamma,$$

con Γ frontiera dell'insieme Ω . Dal punto di vista teorico si richiede

$$R(x, y) \leq 0, \quad (x, y) \in \bar{\Omega} (\equiv \Omega \cup \Gamma).$$

affinchè la soluzione $u(x, y)$ esista e sia unica. Una proprietà che la funzione $u(x, y)$ possiede è quella di max-min.

In dettaglio se $S(x, y) \equiv 0$ e $R(x, y) \leq 0$, allora una soluzione non costante dell'equazione ellittica lineare gode della seguente **proprietà (debole) di max-min**: la funzione $u(x, y)$ non può assumere un massimo positivo o un minimo negativo in Ω .

Se $S(x, y) \equiv 0$ e $R(x, y) \equiv 0$, allora una soluzione non costante dell'equazione ellittica lineare gode della seguente **proprietà (forte) di max-min**: la funzione $u(x, y)$ non può assumere massimo o minimo in Ω .

Dal punto di vista numerico la soluzione approssimata viene trovata sempre nello stesso modo: una volta che il dominio è stato discretizzato le derivate seconde sono approssimate con la consueta formula a tre punti

$$u_{xx}(x_i, y_j) \simeq \frac{u_{i+1,j} - 2u_{ij} + u_{i-1,j}}{h^2}$$

$$u_{yy}(x_i, y_j) \simeq \frac{u_{i,j+1} - 2u_{ij} + u_{i,j-1}}{k^2}$$

mentre le derivate prime possono essere discetizzate, per esempio, usando lo schema alle differenze centrali:

$$u_x(x_i, y_j) \simeq \frac{u_{i+1,j} - u_{i-1,j}}{2h}, \quad u_y(x_i, y_j) \simeq \frac{u_{i,j+1} - u_{i,j-1}}{2k}$$

ottenendo così il seguente schema numerico:

$$\frac{u_{i+1,j} - 2u_{ij} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2u_{ij} + u_{i,j-1}}{k^2} + P_{ij} \frac{u_{i+1,j} - u_{i-1,j}}{2h} +$$

$$Q_{ij} \frac{u_{i,j+1} - u_{i,j-1}}{2k} + R_{ij} u_{ij} + S_{ij} = 0$$

che può essere scritto in forma del tutto equivalente moltiplicando per $2h^2k^2$:

$$2k^2 (u_{i+1,j} - 2u_{ij} + u_{i-1,j}) + 2h^2 (u_{i,j+1} - 2u_{ij} + u_{i,j-1}) +$$

$$+ hk^2 (u_{i+1,j} - u_{i-1,j}) P_{ij} + h^2k (u_{i,j+1} - u_{i,j-1}) Q_{ij} + 2h^2k^2 (R_{ij} u_{ij} + S_{ij}) = 0$$

ottenendo lo schema finale:

$$h^2(2 - kQ_{ij})u_{i,j-1} + k^2(2 - hP_{ij})u_{i-1,j} - 2(2k^2 + 2h^2 - h^2k^2R_{ij})u_{ij} + \\ + k^2(2 + hP_{ij})u_{i+1,j} + h^2(2 + kQ_{ij})u_{i,j+1} + 2h^2k^2S_{ij} = 0.$$

La soluzione numerica deve soddisfare anch'essa la proprietà di max-min. In particolare nel caso del metodo a cinque punti vale il seguente risultato teorico:

Teorema 2.4.1 *Siano $P(x, y)$, $Q(x, y)$ ed $R(x, y)$ tre funzioni continue su $\Omega \cup \Gamma$ e $R(x, y) \leq 0$. Siano M_1 ed M_2 due costanti tali che*

$$|P(x, y)| \leq M_1, \quad |Q(x, y)| \leq M_2.$$

Se $S(x, y) \equiv 0$ ed h, k soddisfano le seguenti disuguaglianze

$$M_1 h \leq 2, \quad M_2 k \leq 2$$

allora la soluzione numerica soddisfa la proprietà (debole) di max-min. Inoltre se $R(x, y) \equiv 0$, allora la soluzione numerica soddisfa la proprietà (forte) di max-min.

È ovvio che se i valori M_1 ed M_2 definiti nell'enunciato del precedente teorema sono molto grandi le restrizioni su h e k sono molto severe e si potrebbe correre il serio rischio di dover utilizzare un numero di punti molto elevato e quindi risolvere un sistema dalle dimensioni ancora più elevate. Una soluzione a questo problema è quella di utilizzare una diversa approssimazione per le derivate prime dell'equazione lineare.

2.4.1 Il metodo Upwind per equazioni ellittiche

L'approssimazione di tipo **Upwind** per le derivate prime è la seguente:

$$u_x(x_i, y_j) = \begin{cases} \frac{u_{i+1,j} - u_{i,j}}{h} & P(x_i, y_j) \geq 0 \\ \frac{u_{i,j} - u_{i-1,j}}{h} & P(x_i, y_j) < 0 \end{cases}$$

$$u_y(x_i, y_j) = \begin{cases} \frac{u_{i,j+1} - u_{i,j}}{k} & Q(x_i, y_j) \geq 0 \\ \frac{u_{i,j} - u_{i,j-1}}{k} & Q(x_i, y_j) < 0. \end{cases}$$

Si utilizzano approssimazioni di ordine più basso ma le approssimazioni numeriche soddisfano le condizioni di max-min senza restrizioni sui passi di discretizzazione h, k .

Capitolo 3

Equazioni paraboliche

3.1 Equazioni di Evoluzione

Le equazioni di evoluzione descrivono fenomeni che variano in funzione del tempo, tra gli altri per esempio fenomeni di onde, termodinamici, di dinamica delle popolazioni. Esse sono sostanzialmente delle equazioni differenziali ordinarie alle quali sono aggiunte le variabili spaziali. L'equazione di evoluzione più comune è la cosiddetta **equazione del calore**:

$$u_t(x, t) = u_{xx}(x, t), \quad 0 \leq x \leq L, t \geq 0$$

con condizione iniziale

$$u(x, 0) = f(x), \quad 0 \leq x \leq L$$

e condizioni al contorno

$$u(0, t) = g_1(t) \quad t \geq 0$$

$$u(L, t) = g_2(t) \quad t \geq 0$$

in cui le funzioni g, φ_0 e φ_1 soddisfano le condizioni di omogeneità:

$$f(0) = g_1(0), \quad f(L) = g_2(0). \quad (3.1)$$

L'equazione del calore descrive fenomeni di diffusione termodinamica, infatti $u(x, t)$ rappresenta, per esempio, l'evoluzione nel tempo della densità di calore di una sbarra termoconduttrice di lunghezza L e spessore trascurabile che

viene sottoposta ad una certa temperatura iniziale (la condizione al contorno $f(x)$) e tale che la temperatura agli estremi sia fissata dalle condizioni poste dalle funzioni $g_1(t)$ e $g_2(t)$. Se dopo l'istante iniziale la sorgente di calore viene tolta allora la densità di calore $u(x, t)$ del punto della barra di ascissa x al tempo t soddisfa l'equazione del calore.

L'equazione di conducibilità del calore può essere generalizzata in molti modi:

1. aggiungendo una variabile spaziale (cioè $u = u(x, y, t)$):

$$u_t(x, y, t) = u_{xx}(x, y, t) + u_{yy}(x, y, t);$$

2. aggiungendo un termine forzante $f(x, t)$:

$$u_t(x, t) = u_{xx}(x, t) + f(x, t);$$

3. aggiungendo un coefficiente di diffusione variabile $a(x)$:

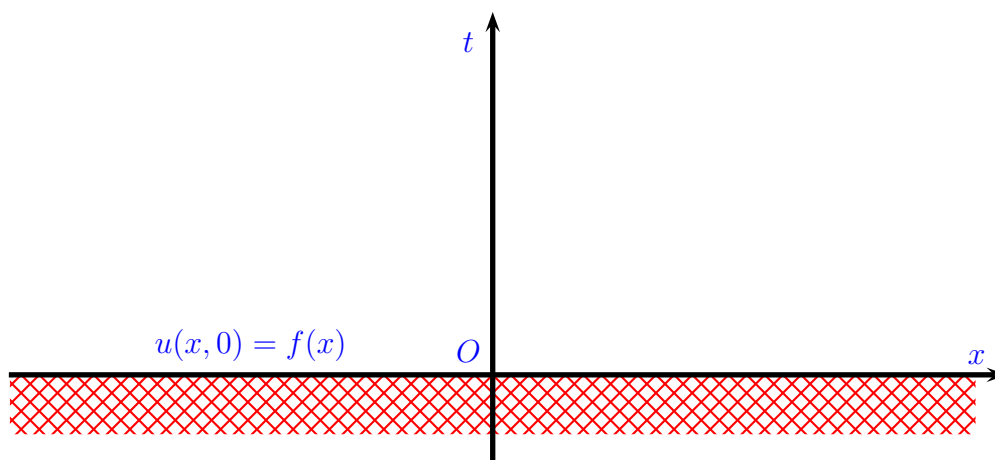
$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x} \left[a(x) \frac{\partial u}{\partial x} \right]$$

con $a(x)$ funzione differenziabile e tale che $0 < a(x) < \infty$ per $x \in [0, L]$;

4. prendendo x appartenente ad un intervallo arbitrario della retta reale e sostituendo le condizioni al contorno con la condizione che la funzione $u(x, t)$ abbia quadrato integrabile, cioè:

$$\int_{-\infty}^{+\infty} [u(x, t)]^2 dx < \infty, \quad t \geq 0.$$

Per l'equazione parabolica è possibile definire due tipi di problemi. Il primo è il problema ai valori iniziali, in cui si tratta di trovare una funzione $u(x, t)$, definita e continua per $x \in \mathbb{R}$ e $t \geq 0$, che soddisfi l'equazione alle derivate parziali per $x \in \mathbb{R}$ e $t > 0$ e la condizione iniziale $u(x, 0) = f(x)$, $x \in \mathbb{R}$, come schematizzato nella seguente figura.



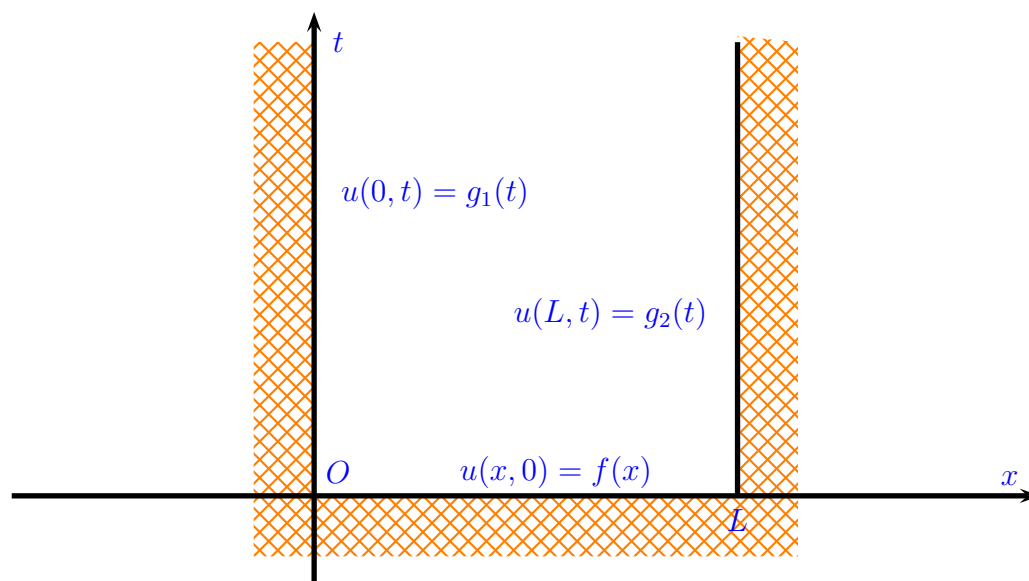
Il secondo è il problema ai valori al contorno, in cui, assegnata una costante $L > 0$, si deve trovare una funzione $u(x, t)$, definita e continua per $0 \leq x \leq L$ e $t \geq 0$, che soddisfi l'equazione alle derivate parziali per $0 < x < L$ e $t > 0$ e le condizioni iniziali:

$$u(x, 0) = f(x) \quad 0 \leq x \leq L$$

$$u(0, t) = g_1(t) \quad t \geq 0$$

$$u(L, t) = g_2(t) \quad t \geq 0.$$

che, a loro volta, soddisfano le condizioni di omogeneità (3.1).



3.2 Il metodo di Eulero per l'equazione del calore

Supponiamo ora di dover risolvere l'equazione

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \quad 0 \leq x \leq L, \quad t \geq 0, \quad (3.2)$$

con una condizione iniziale

$$u(x,0) = f(x), \quad 0 \leq x \leq L \quad (3.3)$$

e condizioni al contorno

$$u(0,t) = g_1(t), \quad u(L,t) = g_2(t), \quad t \geq 0. \quad (3.4)$$

Poiché $t \geq 0$ e non è ovviamente possibile calcolare la soluzione all'infinito si sostituisce $t \geq 0$ con $t \in [0, T_{\max}]$. La costante T_{\max} è generalmente determinata dalla fisica del fenomeno in osservazione. L'equazione viene integrata numericamente in ogni istante di tempo $t_n \leq T_{\max}$. La soluzione numerica di questa equazione, come per tutte le altre equazioni di evoluzione, richiede la discretizzazione dell'equazione stessa rispetto sia al tempo che allo

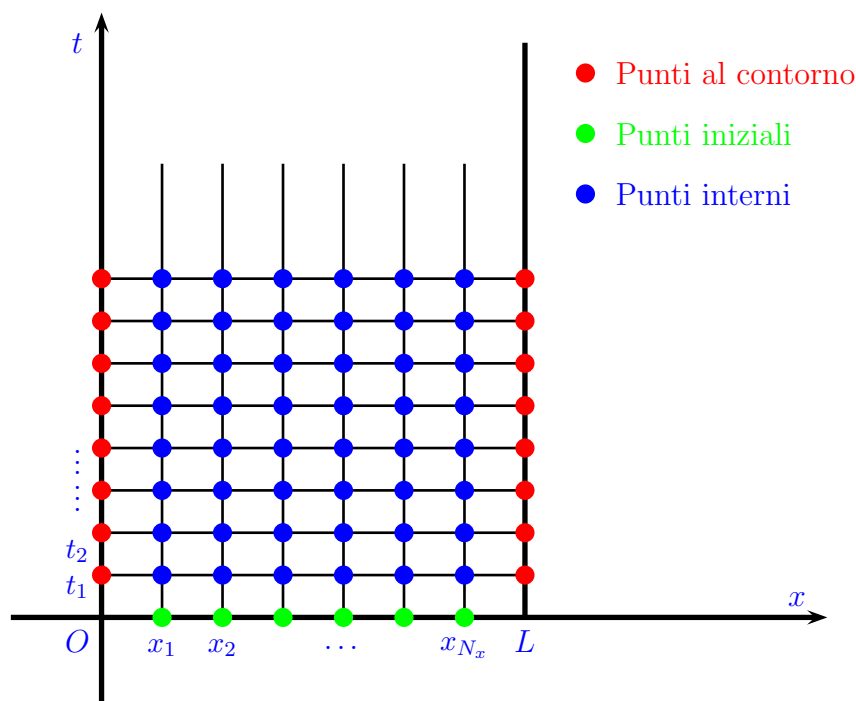
spazio. Il problema discreto è ottenuto mediante l'uso delle differenze finite. Scegliamo un intero positivo N_x e definiamo nella striscia

$$\{(x, t) : x \in [0, L], t \geq 0\}$$

una griglia rettangolare (x_k, t_n) tale che

$$x_k = kh, \quad k = 0, 1, \dots, N_x + 1, \quad t_n = n\Delta t, \quad n \geq 0,$$

essendo $h = L/(N_x + 1)$ e $\Delta t = T_{\max}/N_t$ l'intervallo di tempo tra due approssimazioni successive. I punti (x_i, t_n) del dominio discreto sono di tre tipi, evidenziati nel seguente grafico.



L'approssimazione di $u(x_k, t_n)$ è denotata con $u_{k,n}$. Utilizzando l'operatore differenza centrale possiamo approssimare la derivata parziale seconda nel modo già visto in precedenza:

$$\frac{\partial^2 u}{\partial x^2}(x_k, t_n) \simeq \frac{u_{k+1,n} - 2u_{k,n} + u_{k-1,n}}{h^2}$$

mentre utilizzando l'operatore differenza in avanti per la derivata temporale, si ottiene

$$\frac{\partial u}{\partial t}(x_k, t_n) \simeq \frac{u_{k,n+1} - u_{k,n}}{\Delta t}.$$

Sostituendo in (3.2) si ha

$$\frac{u_{k,n+1} - u_{k,n}}{\Delta t} = \frac{u_{k+1,n} - 2u_{k,n} + u_{k-1,n}}{h^2}$$

$$u_{k,n+1} = u_{k,n} + \frac{\Delta t}{h^2} (u_{k+1,n} - 2u_{k,n} + u_{k-1,n})$$

e infine si ha come risultato il seguente *metodo di Eulero*:

$$u_{k,n+1} = \alpha u_{k-1,n} + (1 - 2\alpha)u_{k,n} + \alpha u_{k+1,n} \quad (3.5)$$

dove $\alpha = \frac{\Delta t}{h^2}$ è una costante che prende il nome di *numero di Courant*.

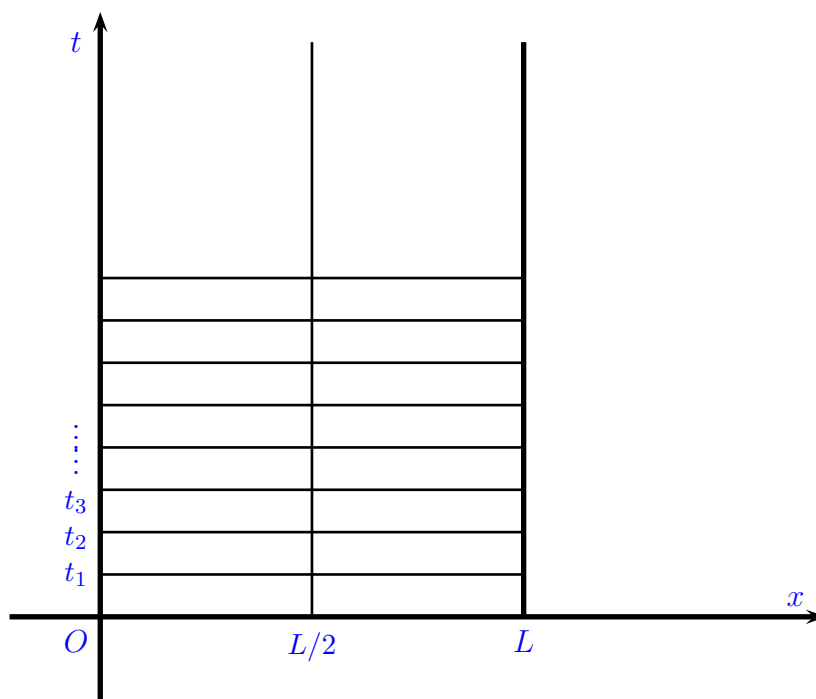
La condizione iniziale (3.3) diventa

$$u_{k,0} = f(kh), \quad k = 1, 2, \dots, N_x.$$

Notiamo che il calcolo di (3.5) per $k = 1$ e $k = N_x$ richiede l'uso delle condizioni sulla frontiera (3.4), che sono

$$u_{0,n} = g_1(t_n) \quad \text{e} \quad u_{N,n} = g_2(t_n).$$

Partendo quindi da $n = 0$ la formula (3.5) consente di determinare esplicitamente le approssimazioni $u_{k,1}$, $k = 1, \dots, N_x$, pertanto è un metodo appunto di tipo esplicito. Tuttavia è noto che i metodi semplici non sono raramente quelli migliori, infatti consideriamo che la soluzione teorica di un'equazione parabolica possiede la stessa proprietà di max-min delle funzioni armoniche è auspicabile che anche la soluzione numerica possieda tale proprietà. Consideriamo la situazione in cui prendiamo un solo nodo interno, cioè consideriamo $h = L/2$ e come condizione iniziale la funzione che assume valore $\varepsilon > 0$ nel punto interno e 0 nei due punti al contorno.



Quindi

$$u_{0,0} = u_{2,0} = 0, \quad u_{1,0} = \varepsilon.$$

Applicando lo schema esplicito appena definito ricaviamo:

$$u_{1,1} = (1 - 2\alpha)u_{1,0} = (1 - 2\alpha)\varepsilon$$

e, ad un generico istante di tempo t_n :

$$u_{1,n} = (1 - 2\alpha)^n \varepsilon.$$

Ora, in base alla proprietà di massimo-minimo deve essere

$$0 \leq u_{1,n} \leq \varepsilon \quad \Rightarrow \quad 0 \leq (1 - 2\alpha)^n \leq 1$$

e quindi

$$0 \leq 1 - 2\alpha \leq 1$$

Poichè $\alpha > 0$ la seconda disuguaglianza è sicuramente verificata, quindi deve essere

$$1 \geq 2\alpha \quad \Rightarrow \quad \alpha \leq \frac{1}{2}.$$

Tale disuguaglianza implica la seguente restrizione sui passi di discretizzazione

$$\Delta t \leq \frac{h^2}{2}.$$

Se $h \simeq 10^{-2}$ deve essere

$$\Delta t \leq 0.5 \cdot 10^{-4}$$

quindi se è necessario integrare l'equazione su intervalli di tempo molto lunghi se deve usare un passo temporale molto piccolo e quindi un numero di passi temporali incredibilmente grande.

3.3 L'Equazione Parabolica Lineare

Consideriamo la generica equazione parabolica lineare

$$u_t(x, t) = P(x, t)u_{xx}(x, t) + Q(x, t)u_x(x, t) + R(x, t)u(x, t) + S(x, t)$$

con $u(x, t)$ soggetta alle condizioni iniziali e al contorno:

$$\begin{aligned} u(x, 0) &= f(x) & 0 \leq x \leq L \\ u(0, t) &= g_1(t) & t \geq 0 \\ u(L, t) &= g_2(t) & t \geq 0. \end{aligned}$$

Per garantire esistenza e unicità della soluzione si deve richiedere che $P(x, t) \geq \nu > 0$ e $R(x, t) \leq 0$. Consideriamo la discretizzazione del dominio già vista nel caso dell'equazione del calore e le seguenti approssimazioni delle derivate parziali:

$$\begin{aligned} u_t(x_i, t_n) &\simeq \frac{u_{i,n+1} - u_{i,n}}{\Delta t} \\ u_x(x_i, t_n) &\simeq \frac{u_{i+1,n} - u_{i-1,n}}{2h} \\ u_{xx}(x_i, t_n) &\simeq \frac{u_{i+1,n} - 2u_{i,n} + u_{i-1,n}}{h^2}, \end{aligned}$$

cosicchè si ottiene il seguente schema

$$\frac{u_{i,n+1} - u_{i,n}}{\Delta t} = P_{in} \frac{u_{i+1,n} - 2u_{i,n} + u_{i-1,n}}{h^2} + Q_{in} \frac{u_{i+1,n} - u_{i-1,n}}{2h} + R_{in}u_{i,n} + S_{in},$$

avendo posto $P_{in} = P(x_i, t_n)$ e analogamente Q_{in}, R_{in} e S_{in} . Il metodo diventa quindi

$$u_{i,n+1} = u_{i,n} + P_{in} \frac{\Delta t}{h^2} (u_{i+1,n} - 2u_{i,n} + u_{i-1,n}) + Q_{in} \frac{\Delta t}{2h} (u_{i+1,n} - u_{i-1,n}) + \Delta t R_{in} u_{i,n} + \Delta t S_{in}.$$

Ponendo

$$\alpha = \frac{\Delta t}{h^2}, \quad \beta = \frac{\Delta t}{2h}$$

si ottiene la forma finale dello schema:

$$u_{i,n+1} = (\alpha P_{in} - \beta Q_{in}) u_{i-1,n} + (1 + \Delta t R_{in} - 2\alpha P_{in}) u_{i,n} + (\alpha P_{in} + \beta Q_{in}) u_{i+1,n} + \Delta t S_{in}.$$

Tale schema, ovviamente esplicito, consente di ottenere la soluzione in tutti i punti x_i allo stesso tempo t_{n+1} , e viene detto **Metodo alle differenze centrali esplicito**. Come abbiamo già accennato la soluzione teorica possiede la proprietà di max-min e, non imponendo alcuna restrizione sui passi di integrazione h e Δt è facile costruire esempi in cui la soluzione numerica viola tale proprietà. Infatti vale il seguente risultato.

Teorema 3.3.1 *Sia $S(x, t) \equiv 0$ e supponiamo che il problema parabolico è definito nel dominio $[0, L] \times [0, T_{\max}]$, con $P(x, t)$, $Q(x, t)$ ed $R(x, t)$ continue (e limitate). Inoltre se per ogni $(x, t) \in [0, L] \times [0, T]$ si ha:*

1. $0 < \nu \leq P(x, t) \leq V$,
2. $|Q(x, t)| \leq M$,
3. $-N \leq R(x, t) \leq 0$,

e

$$Mh \leq 2\nu, \quad \Delta t \leq \frac{h^2}{Nh^2 + 2V}$$

allora la soluzione fornita dal metodo alle differenze centrali esplicito possiede la proprietà di massimo-minimo.

Osserviamo che, poichè $h = L/(N_x + 1)$, dalla prima disequazione segue che

$$\frac{ML}{N_x + 1} \leq 2\nu \Rightarrow N_x + 1 \geq \frac{LM}{2\nu} \Rightarrow N_x > \frac{LM}{2\nu}$$

quindi se M è molto grande (oppure ν è piccolo) il numero di suddivisioni dell'intervallo spaziale deve essere molto elevato.

3.4 Il Metodo Upwind Esplicito

Se vogliamo eliminare tali restrizioni, in particolare quella sul passo h si può utilizzare una diversa discretizzazione per la derivata prima spaziale, usando una diversa approssimazione in base al segno di $Q(x_i, t_n)$:

$$u_x(x_i, t_n) = \begin{cases} \frac{u_{i+1,n} - u_{i,n}}{h} & Q(x_i, t_n) \geq 0 \\ \frac{u_{i,n} - u_{i-1,n}}{h} & Q(x_i, t_n) \leq 0 \end{cases}$$

quindi

$$Q(x_i, t_n)u_x(x_i, t_n) \simeq \frac{(|Q_{in}| - Q_{in})u_{i-1,n} - 2|Q_{in}|u_{i,n} + (Q_{in} + |Q_{in}|)u_{i+1,n}}{2h}.$$

Ponendo

$$\alpha = \frac{\Delta t}{h^2}, \quad \beta = \frac{\Delta t}{2h}$$

si ha il seguente schema:

$$\begin{aligned} u_{i,n+1} = & [\alpha P_{in} + \beta(|Q_{in}| - Q_{in})] u_{i-1,n} \\ & + [1 + \Delta t R_{in} - 2\alpha P_{in} - 2\beta|Q_{in}|] u_{i,n} \\ & + [\alpha P_{in} + \beta(|Q_{in}| + Q_{in})] u_{i+1,n} + \Delta t S_{in}. \end{aligned}$$

Il metodo appena descritto prende il nome di **Metodo Upwind Esplicito** che, rispetto al metodo alle differenze centrali ha il pregio di non richiedere alcuna restrizione sul passo di discretizzazione spaziale. Per ciò che riguarda invece la discretizzazione temporale, nelle stesse ipotesi del Teorema 3.3.1, vale la seguente limitazione:

$$\Delta t \leq \frac{h^2}{Nh^2 + Mh + 2V}.$$

3.5 Metodi numerici per equazioni semilineari

Chiudiamo questo capitolo accennando brevemente alla risoluzione numerica dell'equazione parabolica semilineare:

$$u_t(x, t) = P(x, t)u_{xx}(x, t) + Q(x, t)u_x(x, t) + F(x, t, u) \quad (3.6)$$

tale che, per assicurare l'esistenza e l'unicità della soluzione, la funzione $P(x, t) > 0$ e inoltre F è una funzione limitata e la derivata F_u esiste ed è tale che

$$-M \leq F_u \leq 0, \quad 0 \leq x \leq L, 0 \leq t \leq T_{\max}, u \in \mathbb{R}.$$

I metodi che abbiamo descritto in precedenza si possono applicare anche a questo tipo di equazione, ottenendo così il **Metodo alle differenze centrali**:

$$\begin{aligned} u_{i,n+1} = & (\alpha P_{in} - \beta Q_{in}) u_{i-1,n} + (1 - 2\alpha P_{in}) u_{i,n} + \\ & + (\alpha P_{in} + \beta Q_{in}) u_{i+1,n} + \Delta t F(x_i, t_n, u_{i,n}). \end{aligned}$$

oppure il **Metodo Upwind Esplicito**:

$$\begin{aligned} u_{i,n+1} = & [\alpha P_{in} + \beta(|Q_{in}| - Q_{in})] u_{i-1,n} \\ & + [1 - 2(\alpha P_{in} + \beta|Q_{in}|)] u_{i,n} \\ & + [\alpha P_{in} + \beta(|Q_{in}| + Q_{in})] u_{i+1,n} + \Delta t F(x_i, t_n, u_{i,n}). \end{aligned}$$

Il **Metodo alle differenze centrali implicito** si può ottenere applicando la formula alle differenze all'indietro per la discretizzazione temporale:

$$u_t(x_i, t_n) \simeq \frac{u_{i,n} - u_{i,n-1}}{\Delta t}$$

e applicando la formula alle differenze centrali per la derivata prima spaziale. Lo schema che si ottiene deriva dalla seguente uguaglianza

$$\frac{u_{i,n} - u_{i,n-1}}{\Delta t} = P_{in} \frac{u_{i+1,n} - 2u_{i,n} + u_{i-1,n}}{h^2} + Q_{in} \frac{u_{i+1,n} - u_{i-1,n}}{2h} + F(x_i, t_n, u_{i,n}).$$

Ponendo

$$\alpha = \frac{\Delta t}{h^2}, \quad \beta = \frac{\Delta t}{2h}$$

si arriva al seguente metodo

$$\begin{aligned} & (\alpha P_{in} - \beta Q_{in}) u_{i-1,n} - (1 + 2\alpha P_{in}) u_{i,n} + (\alpha P_{in} + \beta Q_{in}) u_{i+1,n} + \\ & + \Delta t F(x_i, t_n, u_{i,n}) + u_{i,n-1} = 0. \end{aligned}$$

Il metodo è implicito perchè i valori $u_{i-1,n}$, $u_{i,n}$ e $u_{i+1,n}$ sono incogniti e, se la funzione $F(x, t, u)$ non è lineare rispetto a u , ad ogni passo della discretizzazione temporale si deve risolvere un sistema di equazioni non lineari nelle incognite $u_{1,n}, u_{2,n}, \dots, u_{N_x,n}$.

3.6 Il Metodo Upwind Implicito

Abbiamo già introdotto il metodo upwind in forma esplicita. È possibile ottenerne una variante di tipo implicito considerando la seguente approssimazione

$$\frac{u_{i,n} - u_{i,n-1}}{\Delta t} = P_{in} \frac{u_{i+1,n} - 2u_{i,n} + u_{i-1,n}}{h^2} + Q_{in} \frac{u_{i+1,n} - u_{i,n}}{h} + F(x_i, t_n, u_{i,n})$$

se $Q_{in} \geq 0$, e quella alternativa

$$\frac{u_{i,n} - u_{i,n-1}}{\Delta t} = P_{in} \frac{u_{i+1,n} - 2u_{i,n} + u_{i-1,n}}{h^2} + Q_{in} \frac{u_{i,n} - u_{i-1,n}}{h} + F(x_i, t_n, u_{i,n})$$

se $Q_{in} < 0$. Le due formule possono essere sintetizzate nella seguente formulazione:

$$[\alpha P_{in} + \beta(|Q_{in}| - Q_{in})] u_{i-1,n} - [1 + 2(\alpha P_{in} + \beta|Q_{in}|)] u_{i,n} +$$

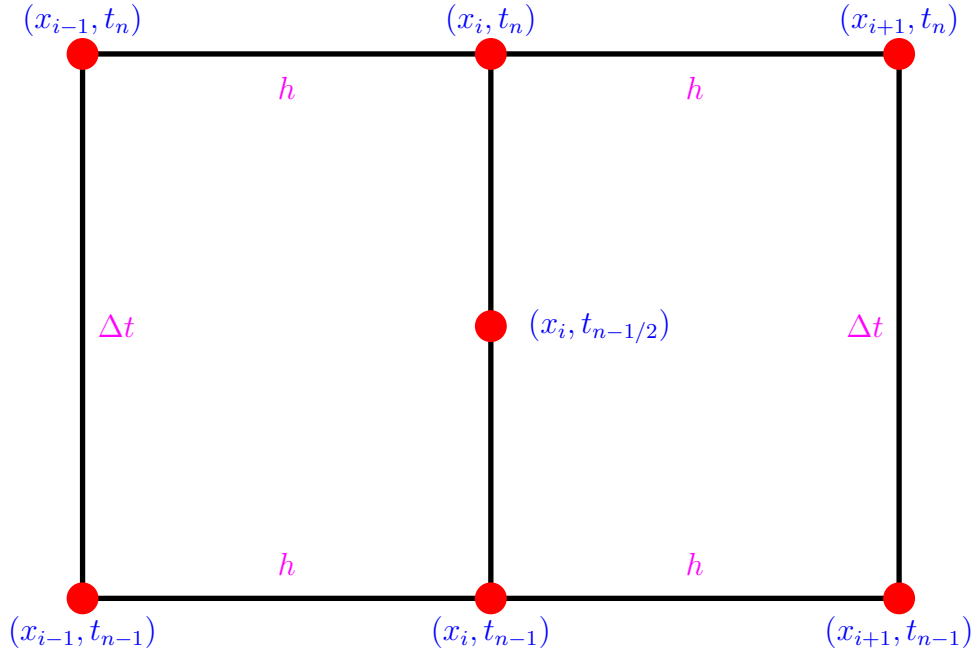
$$[\alpha P_{in} + \beta(|Q_{in}| + Q_{in})] u_{i+1,n} + \Delta t F(x_i, t_n, u_{i,n}) + u_{i,n-1} = 0$$

avendo posto, al solito,

$$\alpha = \frac{\Delta t}{h^2}, \quad \beta = \frac{\Delta t}{2h}.$$

3.6.1 Il Metodo di Crank-Nicolson

Per risolvere numericamente tale equazione si deve osservare innanzitutto che in generale i metodi alle differenze centrali hanno una relativa accuratezza rispetto allo spazio, minore rispetto al tempo. Il motivo di tale comportamento è legato al fatto che le formule più precise sono quelle che usano punti simmetrici rispetto a quello in cui l'approssimazione viene calcolata. Il metodo di Crank-Nicolson utilizza i sei punti $(x_{i\pm 1}, t_n)$, $(x_{i\pm 1}, t_{n-1})$, (x_i, t_n) e (x_i, t_{n-1}) ma imponendo che le approssimazioni soddisfano l'equazione nel punto $(x_i, t_n - \Delta t/2)$.



Poniamo

$$t_{n-1/2} = t_n - (\Delta t)/2$$

e approssimiamo le derivate parziali nel seguente modo:

$$u_t(x_i, t_{n-1/2}) \simeq \frac{u_{i,n} - u_{i,n-1}}{\Delta t}$$

$$u_{xx}(x_i, t_n) \simeq \frac{u_{i+1,n} - 2u_{i,n} + u_{i-1,n}}{h^2}$$

$$u_{xx}(x_i, t_{n-1}) \simeq \frac{u_{i+1,n-1} - 2u_{i,n-1} + u_{i-1,n-1}}{h^2}$$

$$u_{xx}(x_i, t_{n-1/2}) \simeq \frac{1}{2} [u_{xx}(x_i, t_n) + u_{xx}(x_i, t_{n-1})]$$

cosicchè:

$$u_{xx}(x_i, t_{n-1/2}) \simeq \frac{1}{2} \left[\frac{u_{i+1,n} - 2u_{i,n} + u_{i-1,n}}{h^2} + \frac{u_{i+1,n-1} - 2u_{i,n-1} + u_{i-1,n-1}}{h^2} \right].$$

In modo analogo si procede per la derivata prima rispetto a x :

$$u_x(x_i, t_{n-1/2}) \simeq \frac{1}{2} [u_x(x_i, t_n) + u_x(x_i, t_{n-1})].$$

Approssimando le derivate parziali prime con la formula alle differenze centrali risulta:

$$u_x(x_i, t_{n-1/2}) \simeq \frac{1}{2} \left[\frac{u_{i+1,n} - u_{i-1,n}}{2h} + \frac{u_{i+1,n-1} - u_{i-1,n-1}}{2h} \right].$$

Sostituendo le approssimazioni per le derivate nell'equazione (3.6) si ottiene l'espressione del metodo di Crank-Nicolson:

$$\begin{aligned} \frac{u_{i,n} - u_{i,n-1}}{\Delta t} = & P_{i,n-1/2} \left[\frac{u_{i+1,n} - 2u_{i,n} + u_{i-1,n}}{2h^2} + \frac{u_{i+1,n-1} - 2u_{i,n-1} + u_{i-1,n-1}}{2h^2} \right] + \\ & + Q_{i,n-1/2} \left[\frac{u_{i+1,n} - u_{i-1,n}}{4h} + \frac{u_{i+1,n-1} - u_{i-1,n-1}}{4h} \right] + F \left(x_i, t_{n-1/2}, \frac{u_{i,n} + u_{i,n-1}}{2} \right). \end{aligned}$$

Capitolo 4

Equazioni iperboliche

4.1 L'equazione d'onda

Le equazioni iperboliche rappresentano probabilmente la classe che descrive il più ampio numero di fenomeni in diversi campi della fisica e della fluidodinamica. Le equazioni di Eulero per fluidi comprimibili, le equazioni di Einstein per la relatività generale sono esempi di equazioni iperboliche, quasi tutte non lineari. L'equazione lineare di diffusione di tipo iperbolico più nota è sicuramente l'equazione d'onda:

$$u_{xx} - u_{tt} = 0.$$

Ci sono due tipi di equazioni iperboliche:

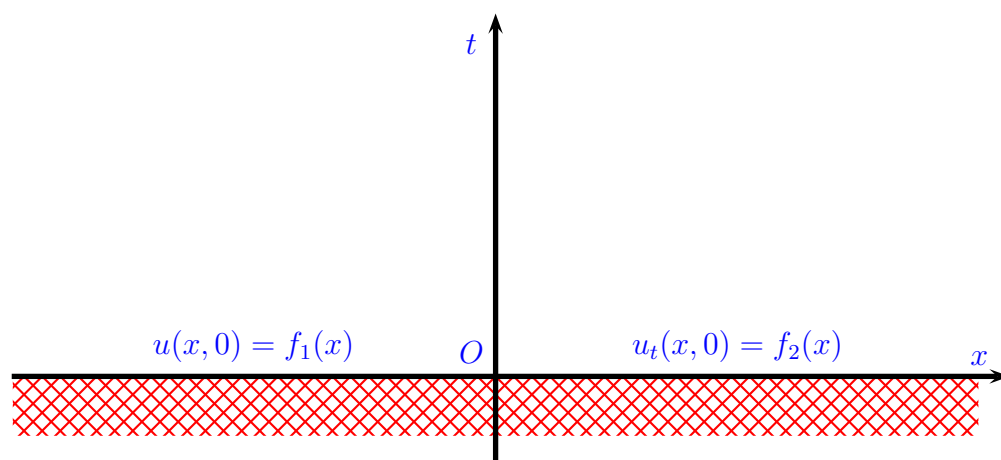
1. Il problema ai valori iniziali (Problema di Cauchy).

Si tratta di trovare una funzione $u(x, t)$, definita e continua per $x \in \mathbb{R}$ e $t \geq 0$, che soddisfi l'equazione alle derivate parziali per $x \in \mathbb{R}$ e $t > 0$ e le condizioni iniziali:

$$u(x, 0) = f_1(x) \quad x \in \mathbb{R}$$

$$u_t(x, 0) = f_2(x) \quad x \in \mathbb{R},$$

come schematizzato nella seguente figura.



2. Il problema ai valori iniziali e al contorno.

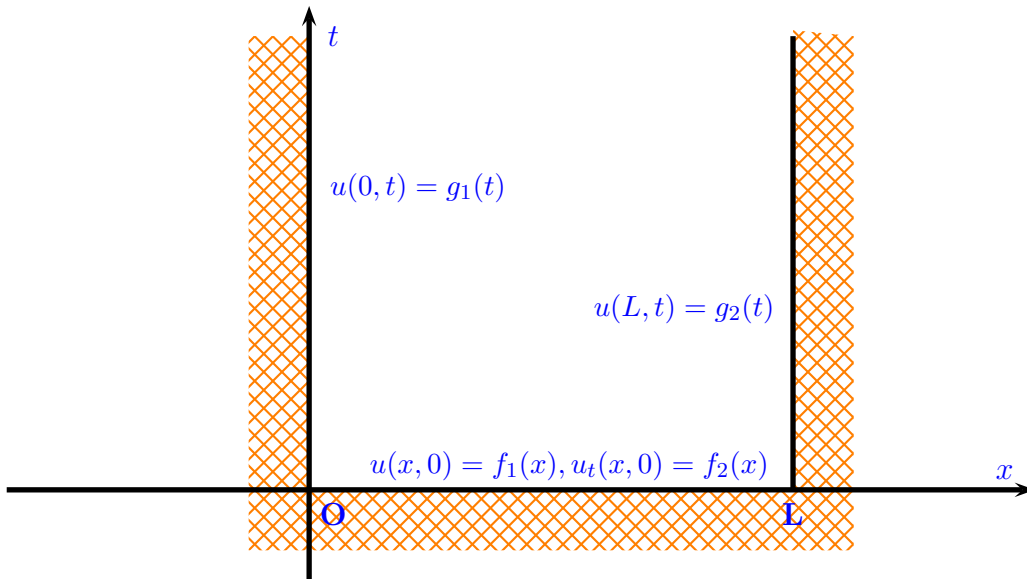
Assegnata una costante $L > 0$ si deve trovare una funzione $u(x, t)$, definita e continua per $0 \leq x \leq L$ e $t \geq 0$, che soddisfi l'equazione alle derivate parziali per $0 < x < L$ e $t > 0$ e le condizioni iniziali:

$$u(x, 0) = f_1(x) \quad 0 \leq x \leq L$$

$$u_t(x, 0) = f_2(x) \quad 0 \leq x \leq L$$

$$u(0, t) = g_1(t) \quad t \geq 0$$

$$u(L, t) = g_2(t) \quad t \geq 0.$$



La risoluzione per via analitica del problema di Cauchy è possibile effettuando il seguente cambio di variabile:

$$\xi = x + t, \quad \psi = x - t$$

e definendo la funzione

$$\mathcal{U}(\xi, \psi) = u(x(\xi, \psi), t(\xi, \psi)) = u\left(\frac{1}{2}(\xi + \psi), \frac{1}{2}(\xi - \psi)\right).$$

Osserviamo innanzitutto che

$$\frac{\partial^2 \mathcal{U}}{\partial \xi \partial \psi} = 0. \tag{4.1}$$

Infatti

$$\frac{\partial \mathcal{U}}{\partial \xi} = \frac{\partial u}{\partial x} \frac{\partial x}{\partial \xi} + \frac{\partial t}{\partial \xi} \frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial u}{\partial x} + \frac{1}{2} \frac{\partial u}{\partial t}$$

e, calcolando la derivata parziale seconda:

$$\begin{aligned} \frac{\partial^2 \mathcal{U}}{\partial \xi \partial \psi} &= \frac{1}{2} \frac{\partial}{\partial \psi} \left[\frac{\partial u}{\partial x} + \frac{\partial u}{\partial t} \right] = \\ &= \frac{1}{2} \left[\frac{\partial x}{\partial \psi} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial x \partial t} \right) + \frac{\partial t}{\partial \psi} \left(\frac{\partial^2 u}{\partial t \partial x} + \frac{\partial^2 u}{\partial t^2} \right) \right] = \\ &= \frac{1}{2} \left[\frac{1}{2} \frac{\partial^2 u}{\partial x^2} + \frac{1}{2} \frac{\partial^2 u}{\partial t \partial x} - \frac{1}{2} \frac{\partial^2 u}{\partial x \partial t} - \frac{1}{2} \frac{\partial^2 u}{\partial t^2} \right] = 0. \end{aligned}$$

L'uguaglianza a zero deriva dall'ipotesi che la funzione $u(x, t)$ soddisfa l'equazione d'onda e dall'uguaglianza delle derivate parziali miste.

Poichè $\mathcal{U}_{\xi\psi} = 0$ possiamo considerare la derivata \mathcal{U}_ξ come funzione della sola variabile ξ quindi integrando (4.1) rispetto a ψ si ottiene:

$$\mathcal{U}_\xi = F_1(\xi)$$

e, integrando nuovamente rispetto a ξ :

$$\mathcal{U}(\xi, \psi) = \int_0^\xi F_1(z) dz + G_2(\psi),$$

dove F_1 e G_2 sono due funzioni arbitrarie differenziabili. Posto

$$G_1(\xi) = \int_0^\xi F_1(z) dz$$

risulta

$$\mathcal{U}(\xi, \psi) = G_1(\xi) + G_2(\psi).$$

Tornando alle variabili x e t si ha che la soluzione deve essere:

$$u(x, t) = G_1(x + t) + G_2(x - t).$$

Sostituendo le condizioni iniziali risulta:

$$u(x, 0) = G_1(x) + G_2(x) = f_1(x)$$

$$u_t(x, 0) = G_1'(x) - G_2'(x) = f_2(x).$$

e, differenziando la prima equazione:

$$G_1'(x) + G_2'(x) = f_1'(x)$$

si ricava agevolmente:

$$G_1'(x) = \frac{1}{2} [f_1'(x) + f_2(x)]$$

$$G_2'(x) = \frac{1}{2} [f_1'(x) - f_2(x)],$$

da cui, integrando rispetto a x , risulta:

$$G_1(x) = \frac{1}{2} \left[f_1(x) + \int_0^x f_2(z) dz \right]$$

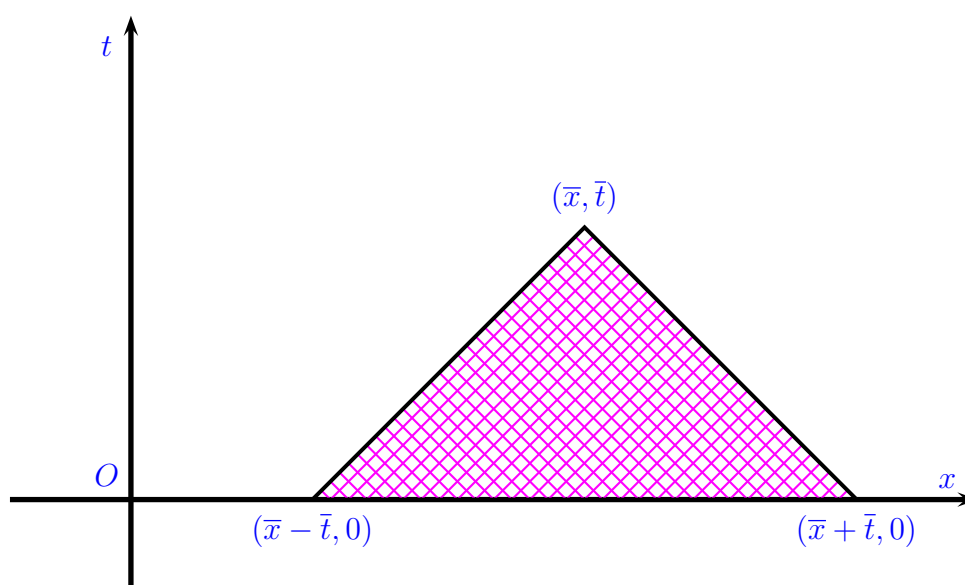
$$G_2(x) = \frac{1}{2} \left[f_1(x) - \int_0^x f_2(z) dz \right].$$

Sostituendo tali formule nell'espressioni di $u(x, t)$ si ottiene:

$$u(x, t) = \frac{1}{2} \left[f_1(x+t) + \int_0^{x+t} f_2(z) dz \right] + \frac{1}{2} \left[f_1(x-t) - \int_0^{x-t} f_2(z) dz \right] =$$

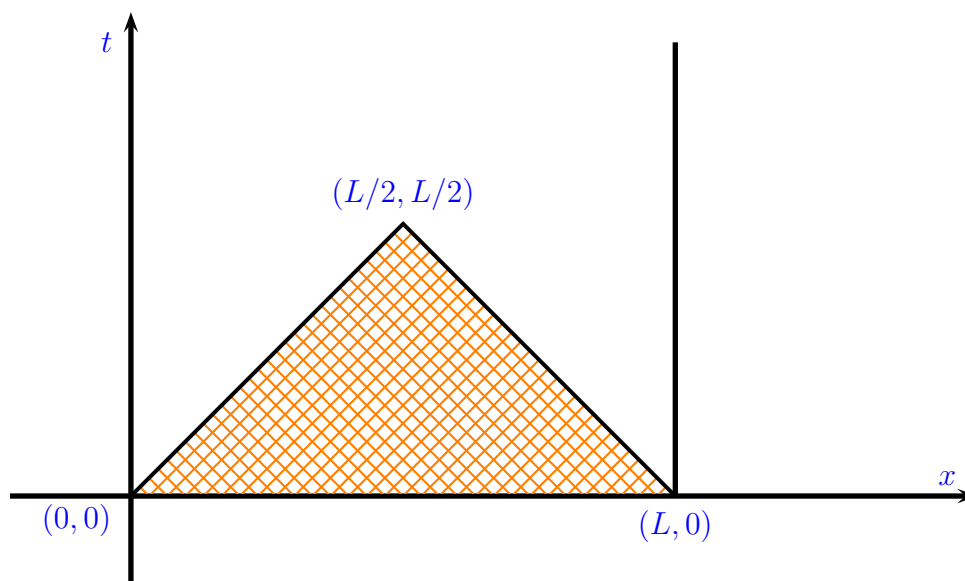
$$= \frac{1}{2} \left[f_1(x+t) + f_1(x-t) + \int_{x-t}^{x+t} f_2(z) dz \right].$$

che prende il nome di **Formula di D'Alembert**. Da tale formula segue che la funzione $u(\bar{x}, \bar{t})$ è determinata univocamente in base alla conoscenza delle funzioni f_1 ed f_2 tra i punti $(\bar{x} - \bar{t}, 0)$ e $(\bar{x} + \bar{t}, 0)$. L'intervallo $[\bar{x} - \bar{t}, \bar{x} + \bar{t}]$ viene detto **intervallo di dipendenza** del punto (\bar{x}, \bar{t}) . La regione interna al triangolo di vertici (\bar{x}, \bar{t}) , $(\bar{x} - \bar{t}, 0)$ e $(\bar{x} + \bar{t}, 0)$ ed evidenziata nella figura seguente si chiama **Regione di dipendenza**.



Le rette congiungenti i punti (\bar{x}, \bar{t}) e $(\bar{x} - \bar{t}, 0)$, (\bar{x}, \bar{t}) e $(\bar{x} + \bar{t}, 0)$ sono dette **caratteristiche** dell'equazione d'onda in (\bar{x}, \bar{t}) .

Osserviamo che, nel caso in cui il problema sia ai valori al contorno, la formula di D'Alembert può essere usata per calcolare la soluzione solo nel triangolo di vertici $(0, 0)$, $(L, 0)$ e $(L/2, L/2)$.



4.2 Un metodo esplicito per l'equazione d'onda

Come al solito si costruisce la griglia suddividendo l'intervallo $[0, L]$ in sottointervalli di ampiezza $h = L/(N + 1)$ e definendo gli istanti di tempo multipli di un valore Δt :

$$x_i = ih, \quad i = 0, 1, 2, \dots, N + 1,$$

$$t_n = n\Delta t, \quad n = 0, 1, 2, \dots$$

Le derivate parziali seconde sono approximate nel modo consueto:

$$u_{xx}(x_i, t_n) \simeq \frac{u_{i+1,n} - 2u_{i,n} + u_{i-1,n}}{h^2}$$

$$u_{tt}(x_i, t_n) \simeq \frac{u_{i,n+1} - 2u_{i,n} + u_{i,n-1}}{(\Delta t)^2}$$

$$\frac{u_{i+1,n} - 2u_{i,n} + u_{i-1,n}}{h^2} - \frac{u_{i,n+1} - 2u_{i,n} + u_{i,n-1}}{(\Delta t)^2} = 0$$

$$u_{i,n+1} - 2u_{i,n} + u_{i,n-1} = \frac{(\Delta t)^2}{h^2}(u_{i+1,n} - 2u_{i,n} + u_{i-1,n})$$

Poniamo $\alpha = (\Delta t)^2/h^2$ e ricaviamo $u_{i,n+1}$:

$$u_{i,n+1} = 2u_{i,n} - u_{i,n-1} + \alpha(u_{i+1,n} - 2u_{i,n} + u_{i-1,n})$$

$$u_{i,n+1} = \alpha u_{i-1,n} + 2(1 - \alpha)u_{i,n} + \alpha u_{i+1,n} - u_{i,n-1}, \quad i = 1, \dots, N, \quad n \geq 1.$$

Il primo insieme di valori che è possibile calcolare è quindi $u_{i,2}$, per i quali è però necessario conoscere $u_{i,1}$, poichè i valori $u_{i,0}$ sono forniti dalla conoscenza della condizione iniziale

$$u_{i,0} = u(x_i, 0) = f_1(x_i).$$

Il problema è ora quello di approssimare la soluzione nei punti $(x_i, \Delta t)$, cioè conoscere i valori:

$$u_{i,1} \simeq u(x_i, \Delta t), \quad i = 1, \dots, N.$$

Per questo motivo si utilizza l'espansione in serie di Taylor:

$$u(x_i, \Delta t) \simeq u(x_i, 0) + \Delta t \frac{\partial u}{\partial t}(x_i, 0) + \frac{(\Delta t)^2}{2} \frac{\partial^2 u}{\partial t^2}(x_i, 0). \quad (4.2)$$

Poichè la funzione $u(x, t)$ soddisfa l'equazione d'onda, allora possiamo sostituire u_{tt} con u_{xx} , e le condizioni iniziali per $u(x, t)$ e $u_t(x, t)$:

$$u(x_i, \Delta t) \simeq f_1(x_i) + \Delta t f_2(x_i) + \frac{(\Delta t)^2}{2} u_{xx}(x_i, 0).$$

L'ultimo termine della serie viene approssimato come al solito:

$$u_{xx}(x_i, \Delta t) \simeq \frac{u(x_{i+1}, 0) - 2u(x_i, 0) + u(x_{i-1}, 0)}{h^2} =$$

$$= \frac{f_1(x_{i+1}) - 2f_1(x_i) + f_1(x_{i-1}))}{h^2}$$

cosicchè si ottiene la seguente approssimazione:

$$u_{i,1} \simeq f_1(x_i) + \Delta t f_2(x_i) + \frac{(\Delta t)^2}{2} \frac{f_1(x_{i+1}) - 2f_1(x_i) + f_1(x_{i-1}))}{h^2}.$$

4.3 Un metodo implicito per l'equazione d'onda

Per risolvere l'equazione d'onda si può discretizzare in modo diverso la derivata seconda di tipo spaziale:

$$\begin{aligned}
 u_{xx}(x_i, t_n) &\simeq \frac{1}{2} [u_{xx}(x_i, t_{n+1}) + u_{xx}(x_i, t_{n-1})] \\
 u_{tt}(x_i, t_n) &\simeq \frac{u_{i,n+1} - 2u_{i,n} + u_{i,n-1}}{(\Delta t)^2} \\
 u_{xx}(x_i, t_{n+1}) &\simeq \frac{u_{i+1,n+1} - 2u_{i,n+1} + u_{i-1,n+1}}{h^2} \\
 u_{xx}(x_i, t_{n-1}) &\simeq \frac{u_{i+1,n-1} - 2u_{i,n-1} + u_{i-1,n-1}}{h^2} \\
 u_{xx}(x_i, t_n) &\simeq \frac{1}{2} \left[\frac{u_{i+1,n+1} - 2u_{i,n+1} + u_{i-1,n+1}}{h^2} + \frac{u_{i+1,n-1} - 2u_{i,n-1} + u_{i-1,n-1}}{h^2} \right].
 \end{aligned}$$

Sostituendo le approssimazioni nell'equazione alle derivate parziali si ottiene:

$$\begin{aligned}
 \frac{u_{i,n+1} - 2u_{i,n} + u_{i,n-1}}{(\Delta t)^2} - \frac{1}{2} \left[\frac{u_{i+1,n+1} - 2u_{i,n+1} + u_{i-1,n+1}}{h^2} + \right. \\
 \left. + \frac{u_{i+1,n-1} - 2u_{i,n-1} + u_{i-1,n-1}}{h^2} \right] = 0.
 \end{aligned}$$

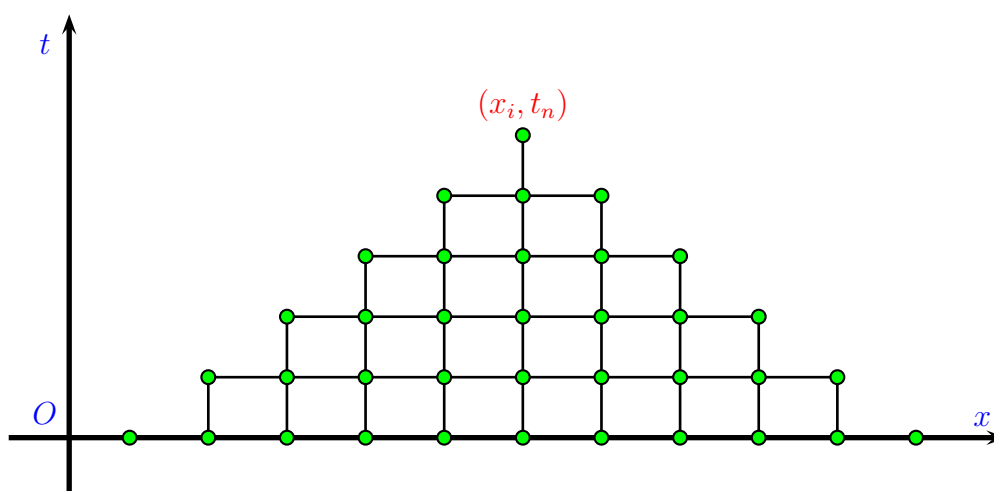
Dopo diversi passaggi si arriva alla formulazione finale:

$$\begin{aligned}
 u_{i-1,n+1} - 2 \left[1 + \left(\frac{h}{\Delta t} \right)^2 \right] u_{i,n+1} + u_{i+1,n+1} = -u_{i-1,n-1} + \\
 -u_{i+1,n-1} - 4 \left(\frac{h}{\Delta t} \right)^2 u_{i,n} + 2 \left[1 + \left(\frac{h}{\Delta t} \right)^2 \right] u_{i,n-1}.
 \end{aligned}$$

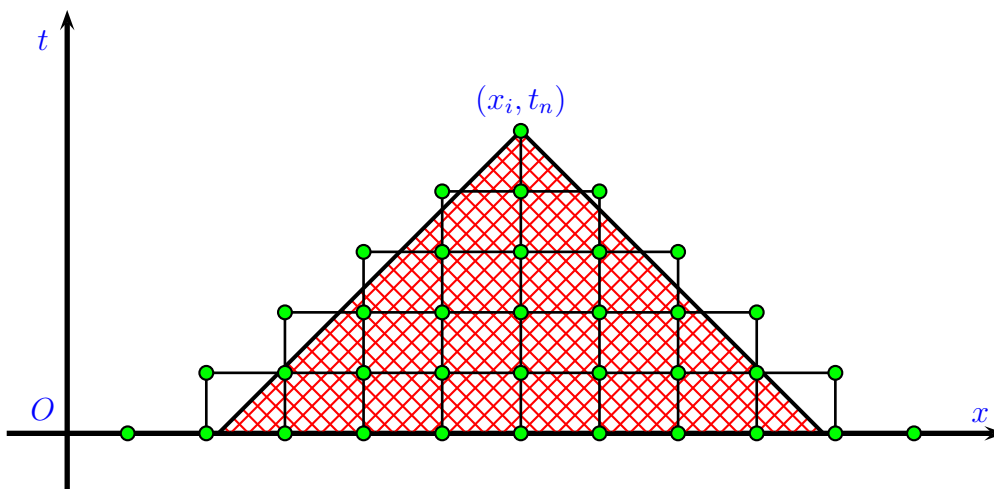
Se la soluzione numerica è nota ai livelli t_n e t_{n-1} allora, utilizzando le condizioni al contorno, la formulazione del metodo costituisce un sistema lineare di $N - 1$ equazioni nelle $N - 1$ incognite $u_{i,n+1}$, $i = 1, N - 1$. Tale sistema ha una struttura tridiagonale a predominanza diagonale quindi ammette un'unica soluzione. Per calcolare la soluzione al livello t_1 si può utilizzare lo stesso di approssimazione (4.2) visto per il metodo esplicito.

4.4 Condizione di Courant, Friedrichs e Levy

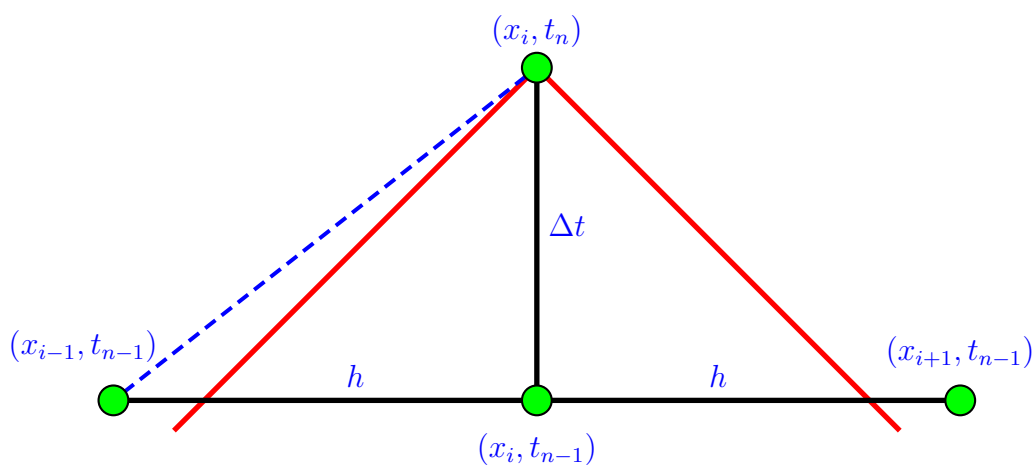
L'approssimazione $u_{i,n}$ dipende, sia nei metodi impliciti che espliciti, da approssimazioni al livello precedente $n - 1$, in particolare da $u_{i,n-1}$, $u_{i-1,n-1}$ e $u_{i+1,n-1}$. A loro volta tali approssimazioni dipendono da altre al livello $n - 2$, in particolare da $u_{i\pm k,n-2}$, con $k = -2, \dots, 2$, così via. In questo modo procedendo a ritroso è possibile definire una specie di dominio di dipendenza discreto che contiene tutte le approssimazioni, dal livello 0 al livello n , necessarie al calcolo di $u_{i,n}$.



Appare ovvio che tale dominio discreto debba avere necessariamente un legame con quello continuo che abbiamo definito in precedenza. Se il dominio continuo includesse quello discreto questo vorrebbe dire che l'approssimazione $u_{i,n+1}$ è stata ottenuta considerando solo una parte dei valori da cui dipende il valore teorico $u(x_i, t_{n+1})$, sicuramente tale approssimazione numerica non può essere un valore affidabile. Al contrario se il dominio discreto contiene quello continuo significa che la soluzione numerica ha utilizzato effettivamente tutti i dati necessari (e anche altri).



Dal punto di vista matematico si deve richiedere che tale situazione si verifichi, imponendo opportune condizioni sui passi di discretizzazione spaziale e temporale. Infatti è necessario richiedere che la retta caratteristica passante per (x_i, t_{n+1}) intersechi la retta di dipendenza del metodo numerico.



La condizione viene verificata se la retta tratteggiata blu ha un coefficiente angolare inferiore rispetto a quello della retta caratteristica (che in questo caso vale 1), cioè se

$$\frac{\Delta t}{h} \leq 1. \tag{4.3}$$

La relazione (4.3) prende il nome di **Condizione di Courant, Friedrichs e Lewy**.

4.5 Il metodo di Lax-Wendroff

Un altro esempio di equazione iperbolica è la seguente

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = 0, \quad a > 0. \quad (4.4)$$

Per verificare che effettivamente è un'equazione iperbolica deriviamola rispetto a t :

$$\frac{\partial^2 u}{\partial t^2} + a \frac{\partial^2 u}{\partial t \partial x} = 0, \quad (4.5)$$

rispetto a x :

$$\frac{\partial^2 u}{\partial t \partial x} + a \frac{\partial^2 u}{\partial x^2} = 0, \quad (4.6)$$

e, sostituendo la derivata mista da (4.5) in (4.6) ricaviamo l'espressione dell'equazione iperbolica del secondo ordine:

$$\frac{\partial^2 u}{\partial t^2} - a^2 \frac{\partial^2 u}{\partial x^2} = 0.$$

Il metodo di Lax-Wendroff, esplicito, risolve numericamente l'equazione (4.4) partendo dall'espansione in serie di Taylor della funzione $u(x_i, t_n + \Delta t)$ rispetto alla variabile temporale e prendendo (x_i, t_n) come punto iniziale:

$$u(x_i, t_n + \Delta t) \simeq u(x_i, t_n) + \Delta t \frac{\partial u}{\partial t}(x_i, t_n) + \frac{(\Delta t)^2}{2} \frac{\partial^2 u}{\partial t^2}(x_i, t_n).$$

Applicando la relazione (4.4):

$$u(x_i, t_n + \Delta t) \simeq u(x_i, t_n) - a \Delta t u_x(x_i, t_n) + \frac{(\Delta t)^2}{2} u_{tt}(x_i, t_n). \quad (4.7)$$

Derivando la (4.4) rispetto a x e rispetto a t si ottiene:

$$u_{tx}(x, t) = -a u_{xx}(x, t), \quad u_{tt}(x, t) = -a u_{xt}(x, t)$$

da cui, applicando nuovamente il teorema di Schwarz, $u_{xt} = u_{tx}$ da cui segue:

$$u_{tt}(x, t) = a^2 u_{xx}(x, t) \quad (4.8)$$

e, sostituendo in (4.7):

$$u(x_i, t_n + \Delta t) \simeq u(x_i, t_n) - a\Delta t u_x(x_i, t_n) + \frac{(a\Delta t)^2}{2} u_{xx}(x_i, t_n). \quad (4.9)$$

Le derivate spaziali vengono approssimate usando la solita formula per $u_{xx}(x_i, t_n)$ e quella alle differenze centrali per la derivata prima.

$$u_{i,n+1} = u_{i,n} - \frac{a\Delta t}{2h} (u_{i+1,n} - u_{i-1,n}) + \frac{(a\Delta t)^2}{2h^2} (u_{i+1,n} - 2u_{i,n} + u_{i-1,n}).$$

Posto

$$\alpha = \frac{a\Delta t}{h}$$

si ottiene lo schema

$$\begin{aligned} u_{i,n+1} &= u_{i,n} - \frac{\alpha}{2} (u_{i+1,n} - u_{i-1,n}) + \frac{\alpha^2}{2} (u_{i+1,n} - 2u_{i,n} + u_{i-1,n}) \\ &= \frac{\alpha}{2} (1 + \alpha) u_{i-1,n} + (1 - \alpha^2) u_{i,n} - \frac{\alpha}{2} (1 - \alpha) u_{i+1,n}. \end{aligned}$$

Capitolo 5

Metodi iterativi per sistemi sparsi

5.1 Il Metodo di Jacobi

Come abbiamo visto nel capitolo dedicato alle equazioni ellittiche i sistemi lineari che derivano dalla discretizzazione di equazioni alle derivate parziali hanno grandi dimensioni e sono sparsi, cioè una buona parte degli elementi della matrice dei coefficienti sono nulli. Tuttavia quando si applicano i metodi diretti a tali sistemi succede che le matrici perdono la struttura di sparsità, cioè molti elementi nulli diventano diversi da zero e inoltre si ha il problema di gestire matrici di grosse dimensioni, il che può causare un notevole degrado delle prestazioni dei metodi usati. Per questi motivi si introduce una nuova classe di metodi, detti *metodi iterativi*. Supponiamo di dover risolvere il sistema $A\mathbf{x} = \mathbf{b}$, con A matrice non singolare, $\mathbf{b} \neq 0$. Assumiamo inoltre che gli elementi diagonali della matrice a_{ii} , $i = 1, \dots, n$, siano diversi da 0. La i -esima equazione del sistema si scrive

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i$$

e, isolando x_i risulta:

$$x_i = \left(b_i - \sum_{j=1, j \neq i}^n a_{ij}x_j \right) \frac{1}{a_{ii}} \quad i = 1, \dots, n.$$

Queste n equazioni sono del tutto equivalenti al sistema di partenza tuttavia la loro forma suggerisce particolari procedimenti iterativi per cercare la

soluzione. A partire da un'approssimazione iniziale $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$ si calcola la successione di vettori $\{\mathbf{x}^{(k)}\}$ ponendo

$$x_i^{(k+1)} = \left(b_i - \sum_{j=1, j \neq i}^n a_{ij} x_j^{(k)} \right) \frac{1}{a_{ii}} \quad i = 1, \dots, n; \quad (5.1)$$

con $k = 0, 1, 2, \dots$.

La generica componente i -esima del vettore al passo $k + 1$ è calcolata per mezzo di tutte le componenti del vettore al passo k eccetto la i -esima. Questo procedimento iterativo prende il nome di *metodo di Jacobi*.

5.2 Il Metodo di Gauss-Seidel

Una variante del metodo di Jacobi si ottiene osservando che, quando si calcola $x_i^{(k+1)}$ si possono utilizzare le approssimazioni $x_j^{(k+1)}$, con $j = 1, \dots, i - 1$, ottenendo

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right). \quad (5.2)$$

Si ottiene in questo modo il classico *Metodo di Gauss-Seidel*, dove la componente i -esima al passo $k + 1$ è calcolata per mezzo delle componenti dalla prima alla $i - 1$ -esima al passo $k + 1$ e dalla $i + 1$ -esima alla n -esima al passo k . Si deve osservare che entrambi i metodi appena introdotti non utilizzano la componente i -esima al passo k . Per questo si introduce una nuova variante che coinvolge tale valore a partire da un parametro $\omega \neq 0$. Si propone lo schema

$$x_i^{(k+1)} = \frac{\omega}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right] + (1 - \omega) x_i^{(k)}. \quad (5.3)$$

Questa classe di metodi prende il nome di *Metodi di Rilassamento*. Si osserva facilmente che se si pone $\omega = 1$ il metodo di Rilassamento coincide con il metodo di Gauss-Seidel.

Tutti i metodi appena descritti tengono in conto il carattere sparso della matrice dei coefficienti poichè si può evitare di eseguire prodotti del tipo $a_{ij} x_j$ quando a_{ij} è nullo. Per decidere quando fermare il calcolo delle iterazioni

si può pensare di fissare a priori una tolleranza ε e prendere $\mathbf{x}^{(k+1)}$ come approssimazione della soluzione quando risulta

$$\frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|}{\|\mathbf{x}^{(k+1)}\|} < \varepsilon$$

per una fissata norma vettoriale. Di solito si sceglie come approssimazione iniziale $\mathbf{x}^{(0)}$ il vettore nullo.

5.3 Convergenza dei metodi iterativi

Per la convergenza dei metodi introdotti conviene dare un'impostazione generale, nota con il nome di teoria dello splitting.

Definizione 5.3.1 Una coppia di matrici M ed N costituiscono uno splitting per la matrice non singolare A se

1. $A = M + N$;
2. $\det M \neq 0$.

Una volta scelto lo splitting, il sistema $A\mathbf{x} = \mathbf{b}$, cioè

$$(M + N)\mathbf{x} = \mathbf{b}$$

si mette nella forma

$$M\mathbf{x} = -N\mathbf{x} + \mathbf{b} \tag{5.4}$$

e, poichè M è non singolare

$$\mathbf{x} = B\mathbf{x} + \mathbf{c} \tag{5.5}$$

avendo posto

$$B = -M^{-1}N, \quad \mathbf{c} = M^{-1}\mathbf{b}. \tag{5.6}$$

L'equazione (5.4) (o la (5.5)) suggeriscono un metodo iterativo per approssimare la soluzione del sistema. Scelto un vettore iniziale $\mathbf{x}^{(0)}$ si costruisce una sequenza

$$M\mathbf{x}^{(k+1)} = -N\mathbf{x}^{(k)} + \mathbf{b}, \quad k = 0, 1, 2, \dots \tag{5.7}$$

o equivalentemente

$$\mathbf{x}^{(k+1)} = B\mathbf{x}^{(k)} + \mathbf{c}, \quad k = 0, 1, 2, \dots \tag{5.8}$$

È evidente che la scelta della matrice M deve essere tale che il sistema lineare (5.7) sia di facile soluzione.

Definizione 5.3.2 *Il metodo iterativo è convergente per un sistema $A\mathbf{x} = \mathbf{b}$ se la successione dei vettori approssimazione è convergente con un qualsiasi vettore iniziale.*

Una definizione equivalente coinvolge la successione dei vettori errore

$$\mathbf{e}^{(k)} = \mathbf{x}^{(k)} - \mathbf{x}, \quad k = 0, 1, 2, \dots \quad (5.9)$$

dove \mathbf{x} è la soluzione teorica del sistema.

Definizione 5.3.3 *Il metodo iterativo è convergente per un sistema $A\mathbf{x} = \mathbf{b}$ se la successione dei vettori errore è infinitesima per ogni vettore errore iniziale.*

Se si sottrae la (5.5) dalla (5.5) si ottiene la cosiddetta equazione degli errori

$$\mathbf{e}^{(k+1)} = B\mathbf{e}^{(k)}, \quad k = 0, 1, 2, \dots$$

Tale equazione è posta anche in una seconda forma che lega $\mathbf{e}^{(k)}$ all'errore iniziale $\mathbf{e}^{(0)}$, infatti, per induzione, è facile verificare che

$$\mathbf{e}^{(k)} = B^k \mathbf{e}^{(0)}.$$

Dalle equazioni degli errori si deduce che la convergenza di un metodo iterativo dipende esclusivamente dallo splitting scelto per la matrice dei coefficienti. La matrice B definita in (5.6) prende il nome di matrice di iterazione del metodo.

Definizione 5.3.4 *Una matrice quadrata B si dice convergente se*

$$\lim_{k \rightarrow +\infty} B^k = 0.$$

Appare banale il seguente risultato.

Teorema 5.3.1 *Un metodo iterativo è convergente se e solo se è convergente la relativa matrice di iterazione.*

Dimostrazione. Dalla seconda equazione degli errori se la matrice è convergente risulta che, per ogni vettore errore iniziale, la successione dei vettori errore tende a zero. Viceversa se

$$\lim_{k \rightarrow +\infty} \mathbf{e}^{(k)} = 0$$

si può scegliere $\mathbf{e}^{(0)} = \mathbf{e}_i$, i -esimo versore di \mathbb{R}^n , e quindi dedurre che ogni colonna di B^k è infinitesima, e quindi B è convergente. \square

Corollario 5.3.1 *Se esiste una norma $\|\cdot\|$ tale che*

$$\|B\| < 1$$

allora il metodo iterativo è convergente.

Dimostrazione. Si ha

$$0 \leq \|B^k\| \leq \|B\|^k,$$

e, per ipotesi, la seconda maggiorazione tende a zero quando k tende a infinito quindi necessariamente anche B^k tende a zero. \square

Corollario 5.3.2 *Se un metodo iterativo è convergente allora il determinante della matrice di iterazione è, in modulo, minore di 1.*

Dimostrazione. Poichè

$$\lim_{k \rightarrow +\infty} B^k = 0$$

anche

$$\lim_{k \rightarrow +\infty} \det B^k = 0$$

e quindi la tesi. \square

Per quello che riguarda gli autovalori di B osserviamo che se λ è autovalore di B allora esiste un vettore non nullo \mathbf{x} tale che

$$B\mathbf{x} = \lambda\mathbf{x}. \tag{5.10}$$

È immediato osservare che λ^m è autovalore di B^m . Infatti moltiplicando (5.10) per B si ha:

$$B^2\mathbf{x} = \lambda B\mathbf{x} = \lambda^2\mathbf{x}$$

e così via generalizzando. Quindi se la matrice B è convergente deve essere

$$\lim_{m \rightarrow \infty} \lambda^m = 0$$

e quindi tutti gli autovalori di B devono essere, in modulo, minori di 1. Ricordiamo che il raggio spettrale della matrice B , $\rho(B)$, è, per definizione il massimo modulo di un autovalore quindi abbiamo la seguente formulazione alternativa del teorema di convergenza.

Teorema 5.3.2 *Un metodo iterativo è convergente se e solo se è il raggio spettrale della matrice di iterazione è minore di 1.*

Vediamo ora come riottenere, dal punto di vista dello splitting i metodi iterativi precedentemente introdotti. Decomponiamo la matrice A come

$$A = L + D + U$$

dove D è la matrice degli elementi diagonali, L è la parte strettamente triangolare inferiore mentre U è la parte strettamente triangolare superiore. Il metodo di Jacobi equivale a scegliere $M = D$ e $N = U + L$, infatti (5.1) si scrive

$$D\mathbf{x}^{(k+1)} = -(L + U)\mathbf{x}^{(k)} + \mathbf{b}, \quad k = 0, 1, 2, \dots$$

che, in termini di ciascuna componente, equivale a

$$a_{ii}x_i^{(k+1)} = b_i - \sum_{j=1, j \neq i}^n a_{ij}x_j^{(k)}$$

che, appunto, coincide con il metodo di Jacobi. Il metodo di Gauss-Seidel corrisponde invece alla scelta $M = D + L$ e $N = U$, infatti (5.2) equivale a

$$(D + L)\mathbf{x}^{(k+1)} = -U\mathbf{x}^{(k)} + \mathbf{b}, \quad k = 0, 1, 2, \dots$$

Per il metodo di Rilassamento abbiamo invece

$$D\mathbf{x}^{(k+1)} = \omega(\mathbf{b} - L\mathbf{x}^{(k+1)} - U\mathbf{x}^{(k)}) + (1 - \omega)D\mathbf{x}^{(k)}$$

da cui

$$\begin{aligned} (D + \omega L)\mathbf{x}^{(k+1)} &= [(1 - \omega)D - \omega U]\mathbf{x}^{(k)} + \omega\mathbf{b} \\ \left(\frac{D}{\omega} + L\right)\mathbf{x}^{(k+1)} &= \left[\left(\frac{1}{\omega} - 1\right)D - U\right]\mathbf{x}^{(k)} + \mathbf{b} \end{aligned}$$

ricavando così

$$M = \frac{D}{\omega} + L, \quad N = \left(1 - \frac{1}{\omega}\right)D + U.$$

Per quello che riguarda l'intervallo di appartenenza del parametro ω vale il seguente risultato.

Teorema 5.3.3 *Condizione necessaria per la convergenza del metodo di Rilassamento è che*

$$|\omega - 1| < 1,$$

quindi se ω è reale:

$$0 < \omega < 2.$$

Dimostrazione. La matrice M è triangolare inferiore e, poichè U ha elementi diagonali nulli è

$$\det M = \det \frac{D}{\omega} = \frac{1}{\omega^n} \det D.$$

Analogamente per la matrice N :

$$\det N = \frac{(\omega - 1)^n}{\omega^n} \det D.$$

Quindi

$$|\det(M^{-1}N)| = |(1 - \omega)^n|.$$

Poiché condizione necessaria per la convergenza di un metodo iterativo è che il modulo del determinante della matrice di iterazione sia minore di 1 abbiamo:

$$|(1 - \omega)^n| < 1 \Rightarrow |1 - \omega| < 1 \Rightarrow 0 < \omega < 2. \quad \square$$

Le condizioni per determinare la convergenza di un metodo iterativo sono spesso di difficile verifica in quanto richiedono la conoscenza di particolari proprietà della matrice di iterazione (per esempio l'insieme degli autovalori) oppure la conoscenza esplicita della stessa matrice, che di solito non si ha. Pertanto dovendo risolvere un sistema lineare $A\mathbf{x} = \mathbf{b}$ si deve applicare il metodo iterativo e sperare che questo converga a meno che ciò non si possa dedurre dalla stessa matrice dei coefficienti A . Infatti vedremo che ci sono alcune classi di matrici per le quali si può, a priori, stabilire la convergenza di un metodo iterativo.

Supponiamo quindi che la matrice A sia a *stretta predominanza diagonale per righe*, cioè

$$\sum_{j=1, j \neq i}^n |a_{ij}| < |a_{ii}|, \quad i = 1, 2, \dots, n \quad (5.11)$$

cioè ogni elemento della diagonale principale è maggiore, in modulo, della somma dei moduli di tutti gli altri elementi della stessa riga. Per la matrice A sia il metodo di Jacobi che quello di Gauss-Seidel convergono. Per il metodo di Jacobi è sufficiente osservare che la matrice di iterazione B può essere

scritta esplicitamente:

$$B = \begin{pmatrix} 0 & -\frac{a_{12}}{a_{11}} & -\frac{a_{13}}{a_{11}} & \cdots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & -\frac{a_{23}}{a_{22}} & \cdots & -\frac{a_{2n}}{a_{22}} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & -\frac{a_{n-1,n}}{a_{n-1,n-1}} \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & \cdots & -\frac{a_{n,n-1}}{a_{nn}} & 0 \end{pmatrix}$$

Considerando la somma dei moduli degli elementi della i -esima riga di B

$$\sum_{j=1, j \neq i}^n \frac{|a_{ij}|}{|a_{ii}|}$$

è ovvio, da (5.11), che è minore di 1 e quindi anche la sua norma infinito (cioè il massimo della somma dei moduli degli elementi di una riga di B) lo è, e pertanto il metodo converge.

Per il metodo di Gauss-Seidel il discorso è ben più complesso. Poichè la matrice è a predominanza diagonale esiste un numero positivo q tale che:

$$\sum_{j=1, j \neq i}^n \frac{|a_{ij}|}{|a_{ii}|} \leq q < 1.$$

Scriviamo la i -esima riga del sistema lineare

$$\sum_{j=1}^{i-1} a_{ij}x_j + a_{ii}x_i + \sum_{j=i+1}^n a_{ij}x_j = b_i \quad (5.12)$$

e, dalla formula del metodo di Gauss-Seidel (5.2), possiamo scrivere

$$\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} + a_{ii}x_i^{(k+1)} + \sum_{j=i+1}^n a_{ij}x_j^{(k)} = b_i. \quad (5.13)$$

Sottraendo (5.12) da (5.13) ricaviamo

$$\sum_{j=1}^{i-1} a_{ij}(x_j^{(k+1)} - x_j) + a_{ii}(x_i^{(k+1)} - x_i) + \sum_{j=i+1}^n a_{ij}(x_j^{(k)} - x_j) = 0$$

Poniamo ora

$$e_j^{(k)} = x_j^{(k)} - x_j, \quad e_j^{(k+1)} = x_j^{(k+1)} - x_j, \quad j = 1, \dots, n$$

e dividiamo per a_{ii} :

$$\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} e_j^{(k+1)} + e_i^{(k+1)} + \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} e_j^{(k)} = 0$$

e ricaviamo $e_i^{(k+1)}$:

$$e_i^{(k+1)} = - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} e_j^{(k+1)} - \sum_{j=i+1}^n \frac{a_{ij}}{a_{ii}} e_j^{(k)}. \quad (5.14)$$

Supponiamo ora che sia p , $1 \leq p \leq n$, l'indice tale che

$$\|\mathbf{e}^{(k+1)}\|_\infty = |e_p^{(k+1)}|.$$

La p -esima equazione di (5.14) si scrive:

$$e_p^{(k+1)} = - \sum_{j=1}^{p-1} \frac{a_{pj}}{a_{pp}} e_j^{(k+1)} - \sum_{j=p+1}^n \frac{a_{pj}}{a_{pp}} e_j^{(k)} \quad (5.15)$$

da cui, passando ai moduli,

$$\begin{aligned} |e_p^{(k+1)}| &\leq \sum_{j=1}^{p-1} \frac{|a_{pj}|}{|a_{pp}|} |e_j^{(k+1)}| + \sum_{j=p+1}^n \frac{|a_{pj}|}{|a_{pp}|} |e_j^{(k)}| \leq \\ &\leq \left(\sum_{j=1}^{p-1} \frac{|a_{pj}|}{|a_{pp}|} \right) \|\mathbf{e}^{(k+1)}\|_\infty + \left(\sum_{j=p+1}^n \frac{|a_{pj}|}{|a_{pp}|} \right) \|\mathbf{e}^{(k)}\|_\infty = \\ &= \alpha \|\mathbf{e}^{(k+1)}\|_\infty + \beta \|\mathbf{e}^{(k)}\|_\infty \end{aligned}$$

dove

$$\alpha = \sum_{j=1}^{p-1} \frac{|a_{pj}|}{|a_{pp}|}, \quad \beta = \sum_{j=p+1}^n \frac{|a_{pj}|}{|a_{pp}|}$$

sono tali che $\alpha + \beta \leq q < 1$. Abbiamo cioè:

$$\|\mathbf{e}^{(k+1)}\|_{\infty} \leq \alpha \|\mathbf{e}^{(k+1)}\|_{\infty} + \beta \|\mathbf{e}^{(k)}\|_{\infty}$$

e quindi

$$\|\mathbf{e}^{(k+1)}\|_{\infty} \leq \frac{\beta}{1-\alpha} \|\mathbf{e}^{(k)}\|_{\infty}.$$

Poniamo

$$\mu = \frac{\beta}{1-\alpha}$$

e proviamo che $\mu \leq q < 1$. Infatti

$$\begin{aligned} \frac{\beta}{1-\alpha} &= \frac{\beta + \alpha - \alpha}{1-\alpha} \leq \frac{q - \alpha}{1-\alpha} = \\ &= \frac{q - \alpha + q\alpha - q\alpha}{1-\alpha} = \frac{q(1-\alpha) - \alpha(1-q)}{1-\alpha} = \\ &= q - \frac{\alpha(1-q)}{1-\alpha} < q, \end{aligned}$$

poichè il secondo addendo dell'ultima equazione è negativo. Quindi

$$\|\mathbf{e}^{(k+1)}\|_{\infty} \leq q \|\mathbf{e}^{(k)}\|_{\infty} \Rightarrow \|\mathbf{e}^{(k)}\|_{\infty} \leq q^k \|\mathbf{e}^{(0)}\|_{\infty} \Rightarrow \mathbf{e}^{(k)} \xrightarrow{k \rightarrow \infty} 0,$$

e quindi la convergenza del metodo. Per le matrici a stretta predominanza diagonale per righe si può dimostrare inoltre che il metodo di Rilassamento converge se $0 < \omega \leq 1$.

Consideriamo ora il caso in cui la matrice A è a stretta predominanza diagonale per colonne, cioè

$$\sum_{i=1, i \neq j}^n |a_{ij}| \leq |a_{jj}|.$$

La matrice di iterazione del metodo di Jacobi B può essere scritta in questo modo

$$B = -D^{-1}(U + L) = -D^{-1}(U + L)D^{-1}D = D^{-1}HD,$$

avendo posto $H = -(U+L)D^{-1}$. Le matrici H e B sono simili, quindi hanno gli stessi autovalori (e lo stesso raggio spettrale). La matrice H è la seguente

$$H = \begin{pmatrix} 0 & -\frac{a_{12}}{a_{22}} & -\frac{a_{13}}{a_{33}} & \cdots & -\frac{a_{1n}}{a_{nn}} \\ -\frac{a_{21}}{a_{11}} & 0 & -\frac{a_{23}}{a_{33}} & \cdots & -\frac{a_{2n}}{a_{nn}} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & -\frac{a_{n-1,n}}{a_{nn}} \\ -\frac{a_{n1}}{a_{11}} & -\frac{a_{n2}}{a_{22}} & \cdots & -\frac{a_{n,n-1}}{a_{n-1,n-1}} & 0 \end{pmatrix}.$$

In questo caso la norma 1 di H (cioè il massimo della somma dei moduli degli elementi delle colonne) è strettamente minore di 1, e anche il raggio spettrale di H (e quindi anche quello di B) lo è ed il metodo converge.

Anche per matrici a stretta predominanza diagonale per colonne anche il metodo di Gauss-Seidel converge (omettiamo la dimostrazione). È importante sottolineare come in questi casi particolari la verifica delle proprietà di dominanza diagonale sia immediata e che inoltre ci sono delle applicazioni in cui è necessario risolvere sistemi lineari di grosse dimensioni e con matrice dei coefficienti sparsa e a predominanza diagonale.

5.4 Il Metodo del Gradiente Coniugato

Consideriamo il problema di minimizzare la funzione

$$\Phi(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} - \mathbf{x}^T \mathbf{b}$$

dove $\mathbf{b} \in \mathbb{R}^n$ e $A \in \mathbb{R}^{n \times n}$ è una matrice simmetrica e definita positiva. La soluzione del problema è:

$$\mathbf{x} = A^{-1} \mathbf{b}$$

quindi ha senso applicare i metodi di minimizzazione di $\Phi(\mathbf{x})$ per risolvere il sistema lineare

$$A \mathbf{x} = \mathbf{b}.$$

Un primo metodo per minimizzare $\Phi(\mathbf{x})$ è il metodo di *steepest descent* (traducibile in italiano come metodo di più ripida discesa). Esso si basa sull'osservazione che nel punto $\mathbf{x}_c \in \mathbb{R}^n$ la funzione Φ decresce più rapidamente nella direzione del gradiente con il segno cambiato:

$$-\nabla\Phi(\mathbf{x}_c) = \mathbf{b} - A\mathbf{x}_c.$$

Poniamo

$$\mathbf{r}_c = \mathbf{b} - A\mathbf{x}_c$$

il vettore detto *residuo* di \mathbf{x}_c . Se il residuo è diverso da zero cerchiamo di trovare il parametro $\alpha > 0$ tale che:

$$\Phi(\mathbf{x}_c + \alpha\mathbf{r}_c) < \Phi(\mathbf{x}_c).$$

Vediamo ora come calcolare il valore di α .

$$\begin{aligned} \Phi(\mathbf{x}_c + \alpha\mathbf{r}_c) &= \frac{1}{2}(\mathbf{x}_c + \alpha\mathbf{r}_c)^T A(\mathbf{x}_c + \alpha\mathbf{r}_c) - (\mathbf{x}_c + \alpha\mathbf{r}_c)^T \mathbf{b} \\ &= \frac{1}{2}(\mathbf{x}_c^T A\mathbf{x}_c + \alpha\mathbf{r}_c^T A\mathbf{x}_c + \alpha\mathbf{x}_c^T A\mathbf{r}_c + \alpha^2\mathbf{r}_c^T A\mathbf{r}_c) - \mathbf{x}_c^T \mathbf{b} - \alpha\mathbf{r}_c^T \mathbf{b} = \\ &= \frac{1}{2}[\alpha^2\mathbf{r}_c^T A\mathbf{r}_c + 2\alpha(\mathbf{r}_c^T A\mathbf{x}_c - \mathbf{r}_c^T \mathbf{b})] + \frac{1}{2}\mathbf{x}_c^T A\mathbf{x}_c - \mathbf{x}_c^T \mathbf{b} \\ &= \frac{1}{2}[\alpha^2\mathbf{r}_c^T A\mathbf{r}_c - 2\alpha\mathbf{r}_c^T \mathbf{r}_c] + \frac{1}{2}\mathbf{x}_c^T A\mathbf{x}_c - \mathbf{x}_c^T \mathbf{b}. \end{aligned}$$

Dovendo minimizzare la funzione $\Phi(\mathbf{x}_c + \alpha\mathbf{r}_c)$ calcoliamo il valore di α che annulla la derivata prima:

$$\Phi'(\mathbf{x}_c + \alpha\mathbf{r}_c) = \alpha\mathbf{r}_c^T A\mathbf{r}_c - \mathbf{r}_c^T \mathbf{r}_c.$$

Quindi

$$\Phi'(\mathbf{x}_c + \alpha\mathbf{r}_c) = 0 \quad \Rightarrow \quad \alpha = \frac{\mathbf{r}_c^T \mathbf{r}_c}{\mathbf{r}_c^T A\mathbf{r}_c}.$$

Abbiamo quindi il seguente algoritmo:

$$\begin{aligned} k &= 0 \\ \mathbf{x}_0 &\text{ arbitrario} \\ \mathbf{r}_0 &= \mathbf{b} - A\mathbf{x}_0 \end{aligned}$$

```

while  $\mathbf{r}_k \neq 0$ 
   $\alpha_k = (\mathbf{r}_k^T \mathbf{r}_k) / (\mathbf{r}_k^T A \mathbf{r}_k)$ 
   $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{r}_k$ 
   $\mathbf{r}_k = \mathbf{b} - A \mathbf{x}_k$ 
   $k = k + 1$ 
end

```

Il vettore \mathbf{x}_k rappresenta un'approssimazione della soluzione. La convergenza del metodo dipende dallo spettro della matrice A ed in particolare dal rapporto tra gli autovalori estremi di A , cioè

$$\frac{\lambda_1(A)}{\lambda_n(A)}.$$

Quando tale rapporto è grande la convergenza è lenta. Inoltre quanto più il vettore \mathbf{x}_k si avvicina alla soluzione tanto più il residuo \mathbf{r}_k ha componenti piccole e quindi la velocità di convergenza tende a diminuire.

Una modifica del metodo consiste nel minimizzare la funzione $\Phi(\mathbf{x})$ lungo una direzione \mathbf{p}_k di decrescita della funzione, cioè tale che

$$\mathbf{p}_k^T \nabla \phi(\mathbf{x}_k) < 0$$

determinando così l'approssimazione

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$$

tale che

$$\phi(\mathbf{x}_{k+1}) = \min_{\alpha \in \mathbb{R}} \Phi(x_k + \alpha_k \mathbf{p}_k).$$

Ripetendo i passaggi in modo già visto si ottiene

$$\Phi'(\mathbf{x}_k + \alpha \mathbf{p}_k) = (\mathbf{x}_k + \alpha \mathbf{p}_k)^T A \mathbf{p}_k - \mathbf{b}^T \mathbf{p}_k$$

da cui, calcolando il punto stazionario, si ottiene

$$\alpha_k = \frac{(\mathbf{b} - A \mathbf{x}_k)^T \mathbf{p}_k}{\mathbf{p}_k^T A \mathbf{p}_k} = \frac{\mathbf{r}_k^T \mathbf{p}_k}{\mathbf{p}_k^T A \mathbf{p}_k}. \quad (5.16)$$

Poichè $\mathbf{r}_k^T \mathbf{p}_k > 0$ segue che $\alpha_k > 0$. Calcolando il residuo al passo $k + 1$ si ottiene l'espressione:

$$\mathbf{b} - A \mathbf{x}_{k+1} = \mathbf{b} - A(\mathbf{x}_k + \alpha_k \mathbf{p}_k) = \mathbf{b} - A \mathbf{x}_k + \alpha_k A \mathbf{p}_k$$

e quindi

$$\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k A \mathbf{p}_k$$

e, dalla (5.16):

$$\mathbf{r}_{k+1}^T \mathbf{p}_k = (\mathbf{r}_k - \alpha_k A \mathbf{p}_k)^T \mathbf{p}_k = \mathbf{r}_k^T \mathbf{p}_k - \alpha_k A \mathbf{p}_k^T \mathbf{p}_k = 0,$$

quindi ad ogni passo il residuo \mathbf{r}_{k+1} è ortogonale al vettore direzione \mathbf{p}_k del passo precedente. Il problema è la scelta delle direzioni \mathbf{p}_k . In particolare si può definire un metodo in cui viene scelta ad ogni passo una direzione \mathbf{p}_k lungo cui muovere le approssimazioni che tenga conto delle direzioni $\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{k-1}$ calcolate ai passi precedenti. Un metodo che utilizza tale strategia è il *metodo del gradiente coniugato* in cui le direzioni sono definite dai seguenti vettori:

$$\mathbf{p}_k = \begin{cases} \mathbf{r}_0 & k = 0 \\ \mathbf{r}_k + \beta_k \mathbf{p}_{k-1} & k \geq 1, \end{cases}$$

dove $\beta_k \in \mathbb{R}$ viene scelto in modo tale che risulti:

$$\mathbf{p}_k^T A \mathbf{p}_{k-1} = 0$$

cioè \mathbf{p}_k risulta A -coniugato rispetto al vettore \mathbf{p}_{k-1} . Imponendo tale proprietà risulta:

$$\begin{aligned} (\mathbf{r}_k + \beta_k \mathbf{p}_{k-1})^T A \mathbf{p}_{k-1} &= 0 \\ \mathbf{r}_k^T A \mathbf{p}_{k-1} + \beta_k \mathbf{p}_{k-1}^T A \mathbf{p}_{k-1} &= 0 \\ \beta_k &= -\frac{\mathbf{r}_k^T A \mathbf{p}_{k-1}}{\mathbf{p}_{k-1}^T A \mathbf{p}_{k-1}}, \quad k \geq 1. \end{aligned} \tag{5.17}$$

La direzione \mathbf{p}_k è una direzione di decrescita di $\Phi(\mathbf{x})$, infatti

$$-\mathbf{p}_k^T \nabla \Phi(\mathbf{x}_k) = p_k^T \mathbf{r}_k = \mathbf{r}_k^T \mathbf{r}_k + \beta_k \mathbf{p}_{k-1}^T \mathbf{r}_k = \mathbf{r}_k^T \mathbf{r}_k > 0.$$

Il valore di α_k viene trovato nello stesso modo descritto nel metodo di più ripida discesa.

$$\begin{aligned}
\Phi(\mathbf{x}_k + \alpha_k \mathbf{p}_k) &= \frac{1}{2}(\mathbf{x}_k + \alpha_k \mathbf{p}_k)^T A(\mathbf{x}_k + \alpha_k \mathbf{p}_k) - (\mathbf{x}_k + \alpha_k \mathbf{p}_k)^T \mathbf{b} \\
&= \frac{1}{2}(\mathbf{x}_k^T A \mathbf{x}_k + \alpha_k \mathbf{p}_k^T A \mathbf{x}_k + \alpha_k \mathbf{x}_k^T A \mathbf{p}_k + \alpha_k^2 \mathbf{p}_k^T A \mathbf{p}_k) - \mathbf{x}_k^T \mathbf{b} - \alpha_k \mathbf{p}_k^T \mathbf{b} \\
&= \frac{1}{2} [\alpha_k^2 \mathbf{p}_k^T A \mathbf{p}_k + 2\alpha_k (\mathbf{p}_k^T A \mathbf{x}_k - \mathbf{p}_k^T \mathbf{b})] + \frac{1}{2} \mathbf{x}_k^T A \mathbf{x}_k - \mathbf{x}_k^T \mathbf{b} \\
&= \frac{1}{2} (\alpha_k^2 \mathbf{p}_k^T A \mathbf{p}_k - 2\alpha_k \mathbf{p}_k^T \mathbf{r}_k) + \frac{1}{2} \mathbf{x}_k^T A \mathbf{x}_k - \mathbf{x}_k^T \mathbf{b}.
\end{aligned}$$

Il valore di α_k che minimizza la funzione è

$$\alpha_k = \frac{\mathbf{p}_k^T \mathbf{r}_k}{\mathbf{p}_k^T A \mathbf{p}_k} = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{p}_k^T A \mathbf{p}_k}.$$

Ora

$$\mathbf{r}_k^T \mathbf{r}_{k-1} = \mathbf{r}_k^T \mathbf{p}_{k-1} - \beta_{k-1} \mathbf{r}_k^T \mathbf{p}_{k-2} = -\beta_{k-1} \mathbf{r}_k^T \mathbf{p}_{k-2}$$

e

$$\mathbf{r}_k^T \mathbf{p}_{k-2} = \mathbf{r}_{k-1}^T \mathbf{p}_{k-2} - \alpha_k \mathbf{p}_{k-1}^T A \mathbf{p}_{k-2} = 0$$

da cui segue che

$$\mathbf{r}_k^T \mathbf{r}_{k-1} = 0, \quad (5.18)$$

cioè ogni vettore residuo è ortogonale al precedente. Inoltre

$$\mathbf{p}_k^T \mathbf{r}_{k-1} = \mathbf{r}_k^T \mathbf{r}_{k-1} + \beta_k \mathbf{p}_{k-1}^T \mathbf{r}_{k-1} = \beta_k \mathbf{p}_{k-1}^T \mathbf{r}_{k-1} = \beta_k \mathbf{r}_{k-1}^T \mathbf{r}_{k-1} \quad (5.19)$$

e

$$\mathbf{p}_k^T \mathbf{r}_{k-1} = \mathbf{p}_k^T \mathbf{r}_k + \alpha_{k-1} \mathbf{p}_k^T A \mathbf{p}_{k-1} = \mathbf{p}_k^T \mathbf{r}_k = \mathbf{r}_k^T \mathbf{r}_k \quad (5.20)$$

ottenendo una nuova espressione per β_k :

$$\beta_k = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{r}_{k-1}^T \mathbf{r}_{k-1}}.$$

Possiamo riassumere il metodo del gradiente coniugato nel seguente algoritmo:

```
 $k = 0$   
 $\mathbf{x}_0$  arbitrario  
 $\mathbf{r}_0 = b - A\mathbf{x}_0$   
while  $\mathbf{r}_k \neq 0$   
     $\beta_k = (\mathbf{r}_k^T \mathbf{r}_k) / (\mathbf{r}_{k-1}^T \mathbf{r}_{k-1})$  ( $\beta_0 = 0$  se  $k = 0$ )  
     $\mathbf{p}_k = \mathbf{r}_k + \beta_k \mathbf{p}_{k-1}$  ( $\mathbf{p}_0 = \mathbf{r}_0$  se  $k = 0$ )  
     $\alpha_k = (\mathbf{r}_k^T \mathbf{r}_k) / (\mathbf{p}_k^T A \mathbf{p}_k)$   
     $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$   
     $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k A \mathbf{p}_k$   
     $k = k + 1$   
end
```